https://crad.ict.ac.cn
63(2): 393−404, 2026

基于图的元学习的小样本虚假评论检测算法

陈 卓 狄家奇 张 傲 卢晓鹏 杜军威
（青岛科技大学数据科学学院 山东青岛 266061）
（15704626211@163.com）

Graph-Based Meta-Learning Algorithm for Few-Shot Fake Review Detection

Chen Zhuo, Di Jiaqi, Zhang Ao, Lu Xiaopeng, and Du Junwei
（*School of Data Science*, *Qingdao University of Science and Technology*, *Qingdao*, *Shandong* 266061）

Abstract Graph-based fake review detection primarily faces the notoriously challenge of effectively aggregating neighbor information from different relations particularly in an attributed graph with only a limited number of labeled positive samples. The goal is to enhance the sensitivity of graph representation learning to anomalous nodes and improve its generalization capability. To address this challenge, we propose a meta-learning-based multi-information fusion graph difference network (Meta-MGDN). By innovatively constructing a multi-view partitioning and multi-information fusion module, the structural and attribute information of users, items, and rating relationships are fully explored to enable the network to capture multi-faceted information and uncover and analyze relationships between review nodes. We design a multi-view neighborhood difference aggregation module that integrates neighborhood information with self-neighborhood difference information, enabling the network to simultaneously capture both the correlation and the differences between nodes, thereby significantly enhancing its sensitivity to potentially anomalous review nodes. Finally, a meta-learning framework is introduced to simulate the few-shot problem using multiple auxiliary networks, which helps the network practically maintain a high generalization ability in the critical issue of few-shot fake review detection scenarios. Experiments on real-world datasets show that Meta-MGDN significantly outperforms the state-of-the-art baselines in the field of graph-based fake review detection.

Key words fake review detection; graph neural network; meta-learning; few-shot; graph anomaly detection

摘 要 基于图的虚假评论检测主要面临着如何在仅有少量正样本标注的属性图中，有效聚合图中不同关系的邻居信息，提高图表示学习对于异常节点的敏感性和泛化能力的挑战。针对此挑战，提出基于元学习的多信息融合图差异网络（multi-information fusion graph deviation network based on meta-learning, Meta-MGDN）。通过构建多视图划分与多信息融合模块，充分挖掘用户、项目、评分的结构信息与属性信息，以实现网络对多方面信息的获取并挖掘评论节点之间的关系。设计多视图邻域差异聚合模块，合并邻域信息与自身-邻域差异信息，使网络同时关注节点之间的关联性与差异性，提高了网络对于异常节点的敏感性。最后，引入元学习框架，利用多个辅助网络增强目标网络学习小样本任务的经验，从而在小样本虚假评论检测场景下保持较高泛化能力。在真实公开评论数据集上进行的实验表明，Meta-MGDN 在基于图的虚假评论检测领域上的效果优于先进的基线。

关键词 虚假评论检测；图神经网络；元学习；小样本；图异常检测

中图法分类号 TP399

DOI: 10.7544/issn1000-1239.202440717 **CSTR:** 32373.14.issn1000-1239.202440717

收稿日期：2024-09-01；**修回日期**：2025-07-22

在数字购物时代,评论是消费者决策的关键,可帮助买家规避风险,为厂商提供改进产品和服务的宝贵意见。优质评论能促进消费者间的信息共享,增强在线购物社区的价值。然而,一些恶意商家或用户出于营销目的和利益驱使,会发表虚假评论,造成对消费者和市场的负面影响。研究发现,一些知名网站如 Proceline, Orbitz, Tripadivisor, Expedia 上的虚假评论比例为 2%~6%[1-2];一些网评平台如 Yelp,虚假评论比例可达 7%~10%。虚假评论会使消费者植入错误的预期,做购买决策时受到误导,影响消费满意度;会使商家受到不公平的竞争压力;也会削弱公众对网络信息真实性的信任,影响社会诚信体系的稳定性[3]。因此,研究如何有效检测虚假评论对于保护消费者权益、维护商家信誉、促进电子商务市场健康发展以及推动技术进步都具有重要意义。

传统虚假评论检测方法依赖于特征工程和机器学习算法,需要大量人工参与,不仅耗时且容易受到主观判断的影响[4]。为了克服此问题,研究者们应用深度学习的方法,从文本数据中自动学习特征表示,减少人为干预[5]。但此类方法主要关注于文本内容,往往忽略了评论、用户和商品之间的复杂交互关系。为了更全面地捕捉这些交互关系,研究者们开始探索使用图网络模型进行虚假评论检测。基于图的虚假评论检测的优势在于上下文建模能力和领域信息聚合能力[6]。图结构便于建模评论、用户和商品节点间的复杂交互关系,通过邻域聚合机制融合相邻节点的多维特征,捕获节点间的相互影响和潜在关系,提高虚假评论检测的准确性。

目前基于图的虚假评论检测方法通常将评论、用户、商品 3 类异质节点映射到同一特征空间进行邻域聚合,然而这种方式易导致信息混淆的问题[7]。此外,邻域聚合虽然关注节点之间的关联性,但忽视了节点间的个体差异性。由于虚假评论检测不满足同质性假设,因此,如果将邻域内正常节点特征聚合到虚假评论节点上,会抑制虚假评论特征的特异性。再者,现有研究大多采用监督学习方法,依赖大量标注数据学习图节点表示。然而真实场景中,对虚假评论的识别和标注困难,可用于模型训练的有效样本较为稀缺,因而限制了模型的学习效果[8]。

本文针对以上问题,提出基于元学习的多信息融合图差异网络(multi-information fusion graph deviation network based on meta-learning, Meta-MGDN)。首先,该模型设计了多视图划分与多信息融合模块和多视图邻域差异聚合模块,以进行更精准的图表示学习。接着,引入元学习框架,借助辅助网络进行小样本学习,赋予模型较强的泛化能力。具体而言,本文的工作总结如下:

1) 设计多视图划分与多信息融合模块,将评论图划分为多个评论视图,并将多种信息融入到评论节点中,以实现网络对多方面信息的获取,进行更好的节点表示。

2) 提出多视图邻域差异聚合模块,用于合并邻域信息和自身-邻域差异信息,网络能够同时关注节点之间的关联性和差异性,从而提高对虚假评论的敏感性。

3) 引入元学习模块,通过多个辅助网络来模拟小样本问题对网络进行训练,然后利用目标网络进行微调,使得目标网络在小样本场景下具有更好的表现。

4) 在真实公开评论数据集中进行广泛的实验,验证算法检测虚假评论节点的有效性。

1 相关工作

下面从基于图的虚假评论检测方法和图元学习方法以及图神经网络图异常检测 3 方面介绍本文的相关工作。

1.1 基于图的虚假评论检测方法

基于图的虚假评论检测利用了图结构的多种特性,能够克服过去机器学习方法以及部分深度学习方法存在的多个缺陷。Alhosseini 等人[9]提出图卷积网络模型,通过采用归纳表示学习方法,利用评论者的个人资料信息构建社交网络图进行虚假评论检测。Yang 等人[10]提出一种半监督训练的图网络模型,从评论日志中提取文本和行为特征,并将其转换为基于规则的图结构进行虚假评论检测。Ghadery 等人[11]提出一种深度卷积网络模型,其中包含 3 个不同的目标函数,采用并行卷积层,从评论中提取文本特征进行虚假评论检测。Wu 等人[12]结合异质和同质图方法,捕捉评论、用户和产品之间的局部上下文来检测虚假评论。Yang 等人[13]在食品的虚假评论检测领域引入图神经网络架构,通过设计分层注意力机制来同时捕捉用户的静态特征和评论行为的动态特征。

1.2 图的元学习方法

近年来,元学习作为解决小样本问题的方法,受到广泛关注。它通过跨任务学习来快速适应新任务,并在小样本情况下拥有良好的泛化能力。图元学习

专注于学习图结构和属性，通过增强模型的学习和推断能力，提升图网络处理小样本的效果。Huang 等人[14]使用多个卷积层并结合注意力模块，为每个节点构建元图来实现少样本分类。Yao 等人[15]利用从辅助图中迁移学习的知识，提高新目标图上的半监督节点分类效果。Ding 等人[16]通过局部子图传递子图特有信息，并通过元梯度更快地学习可转移的信息，以完成节点分类任务。Cao 等人[17]在元学习框架上只根据少量相关的三元组预测一个关系的新三元组，来实现小样本的链接预测。Ding 等人[18]构建多网络元学习框架来检测图节点异常。本文重点研究运用图元学习框架来缓解虚假评论检测中小样本的问题。

1.3 图神经网络

图神经网络是一种将图数据映射到特征空间的方法。最初受到图谱理论的启发，基于谱的图卷积网络应运而生，并通过设计不同的图卷积层，展示出其在实际应用中的有效性[19]。Kipf 等人[20]提出的采用线性滤波器的图卷积网络模型，已成为最流行的图神经网络模型之一。除了基于谱的图卷积模型外，基于空间的图神经网络也受到了广泛关注。这类方法不为每个节点训练单独的嵌入，而是学习一组聚合函数来从节点的局部邻域中聚合特征。Veličković 等人[21]提出 GraphSAGE 模型，通过学习一个可以推广到未知节点的嵌入函数，实现对网络结构化数据的归纳表示学习。类似地，Hamilton 等人[22]提出 GAT 模型，在聚合节点邻域信息时引入自注意机制来学习节点表示。

1.4 图异常检测

图异常检测（graph anomaly detection, GAD）是指图结构数据中利用各类技术识别结构或属性上显著异常的节点、边或子图，并广泛用于欺诈检测、虚假账号识别等领域。He 等人[23]提出 ADA-GAD 模型，通过设计一种无需学习参数的异常去噪增强方法，生成异常水平降低的图数据，使图自编码器在预训练时更专注于捕捉正常模式。Roy 等人[24]提出了 GAD-NR 模型，将邻域重建作为 GAD 的有效度量，并使用高斯近似邻居特征分布计算重建损失，提高计算效率。Cai 等人[25]将异常生成与知识蒸馏相结合，提出了基于自增强知识蒸馏的联邦图异常检测框架 FGAD，通过对抗生成机制增强异常检测的鲁棒性。

本文聚焦于虚假评论检测任务，利用多视图建模、差异聚合和少样本学习等手段提升了图卷积网络对异常节点的识别能力。

2 基于图的元学习的小样本虚假评论检测

2.1 相关定义

在本节中，首先给出评论图结构的相关定义，然后给出基于元学习的评论图数据划分的描述定义。

定义 1. 评论图。评论图是一种特殊的异质信息网络，包含多种类型的对象和多种类型的连接。其以评论、用户、项目、评分构造为节点，以用户-评论，项目-评论，评分-评论的交互关系构造为边，用$G=(V,\varepsilon,\boldsymbol{A},X,y)$表示。其中$V$是对象集合，包含图中的所有节点；$\varepsilon$是连接集合，表示图中的连接关系；$\boldsymbol{A}$是邻接矩阵，表示图中节点间的连接关系；$X$是属性集合，包含节点的属性信息；$y$是评论节点的真实标签。

定义 2. 基于元学习的评论图数据划分。一般而言，元学习虚假评论检测旨在通过从辅助网络的训练中学习共性的知识和模式，以最大程度提高目标网络上的检测性能。除了目标网络G^{t}外，假设存在P个辅助网络，它们与目标网络G^{t}属于相似的领域。对于网络来说，V^{L}是标签为虚假的节点集合，V^{U}是未标注的评论节点集合。在本文的问题中$|V^{\mathrm{L}}|\ll|V^{\mathrm{U}}|$，因为只提供极少数的有标注数据。定义中基于元学习的评论图数据划分如下所示。

已知：1）P个辅助网络$G^{\mathrm{S}}=\{G_1^{\mathrm{S}},G_2^{\mathrm{S}},\cdots,G_P^{\mathrm{S}}\}$其中$G_i^{\mathrm{S}}=(V_i^{\mathrm{S}},\varepsilon_i^{\mathrm{S}},\boldsymbol{A}_i^{\mathrm{S}},X_i^{\mathrm{S}},y_i)$。2）1 个目标网络$G^{\mathrm{t}}=(V^{\mathrm{t}},\varepsilon^{\mathrm{t}},\boldsymbol{A}^{\mathrm{t}},X^{\mathrm{t}},y^{\mathrm{t}})$。3）每个网络中都包含少量标记虚假评论节点集和未标记的评论节点集$\{V_1^{\mathrm{L}},V_1^{\mathrm{U}}\},\{V_2^{\mathrm{L}},V_2^{\mathrm{U}}\},\cdots,\{V_P^{\mathrm{L}},V_P^{\mathrm{U}}\}$和$\{V_t^{\mathrm{L}},V_t^{\mathrm{U}}\}$。

学习虚假评论检测模型的目标是能够利用多个辅助网络（如$G^{\mathrm{S}}=\{G_1^{\mathrm{S}},G_2^{\mathrm{S}},\cdots,G_P^{\mathrm{S}}\}$）中的虚假信息，来检测目标网络$G^{\mathrm{t}}$中的虚假评论节点。理想情况下，被检测出的虚假节点应具有比正常节点更高的排名分数。

2.2 基于元学习的多信息融合图差异网络

本文采用基于跨网络元学习的框架，提出的 Meta-MGDN 模型主要包括多信息融合图差异网络与元学习 2 个部分。模型整体架构如图 1 所示。

在多信息融合图差异网络部分，旨在充分利用异构节点信息，使网络获取到更加全面、更加具有区别性的信息，并学习欺诈节点与正常节点之间的显著差异性，使网络对异常节点具有敏感性。本网络由 3 个部分组成：多视图划分与多信息融合模块、多视图邻域差异聚合模块和节点评分器。模型架构如图 2 所示。

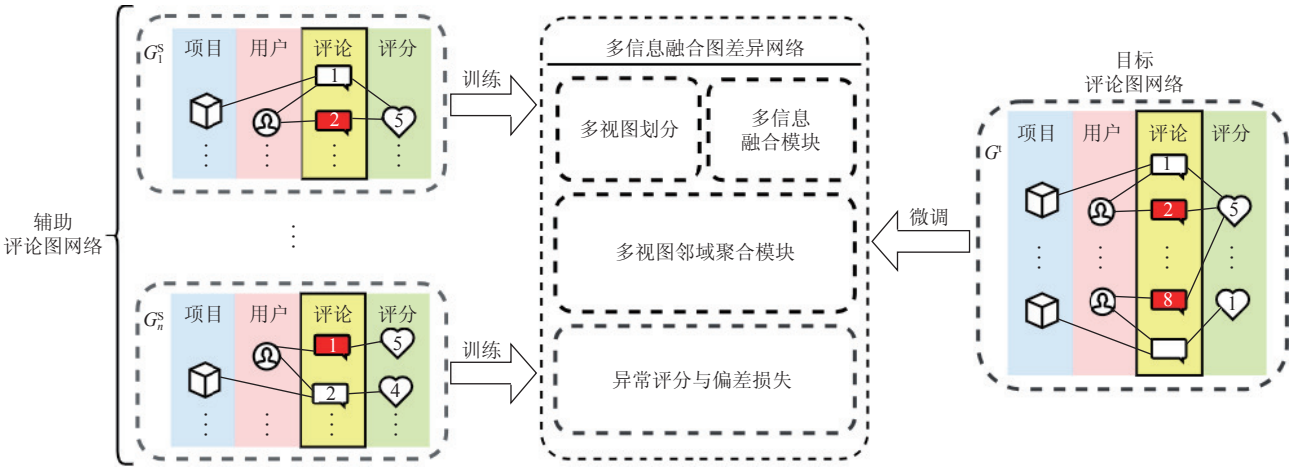


Fig. 1 Overall architecture of Meta-MGDN

图 1 Meta-MGDN 整体架构

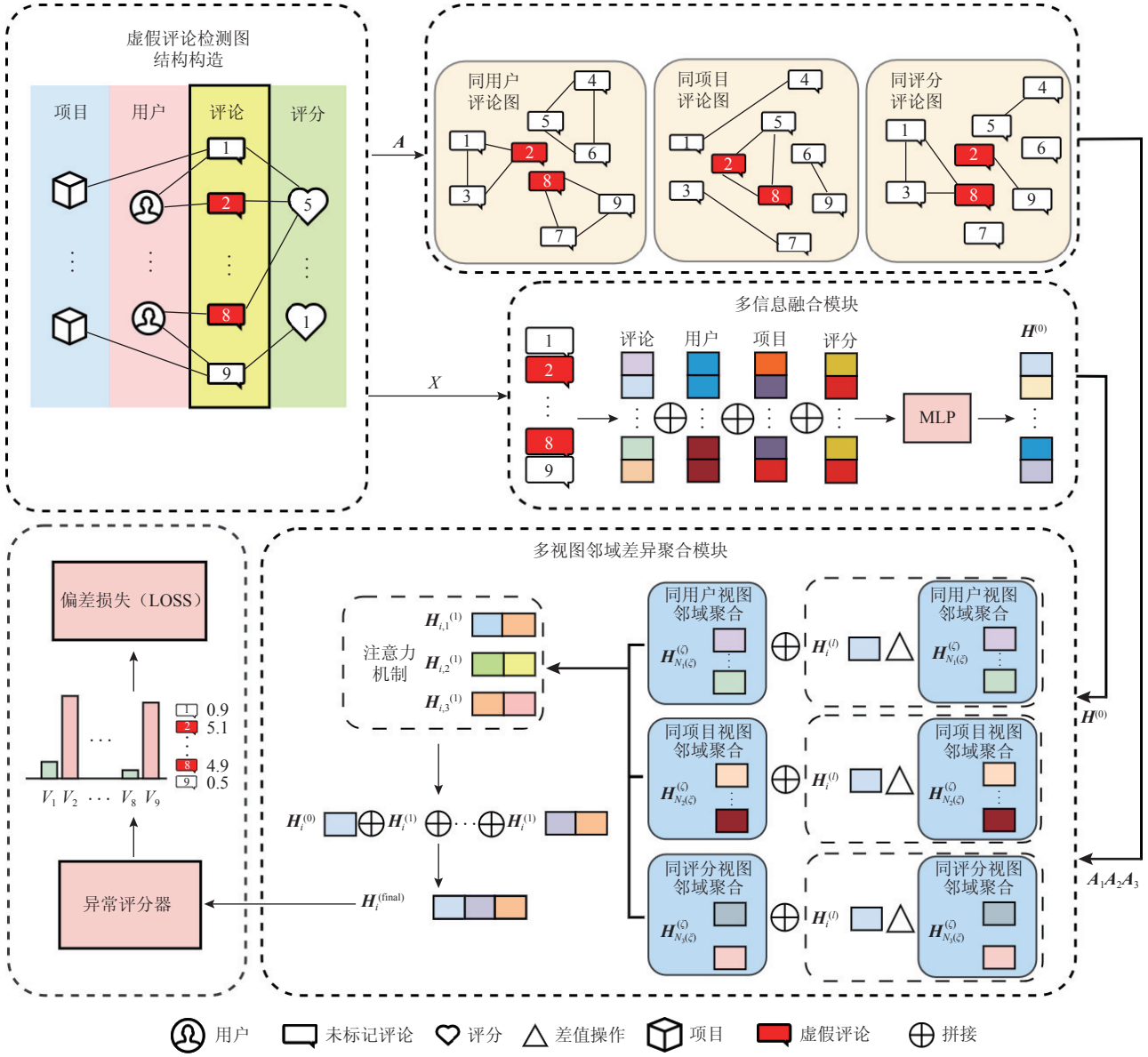


Fig. 2 Architecture of multi-information fusion graph difference network

图 2 多信息融合图差异网络架构

在元学习框架的部分，模型充分利用有限的少量标签使其在小样本的场景下仍能保持较高的泛化性，以及使用多个辅助图网络数据和一个目标图网络数据进行训练。

2.2.1 多视图划分与多信息融合模块

该模块根据评论图中的结构信息划分为多个评论视图，根据用户、项目、评分的属性信息进行多信息融合，以实现网络对多方面信息的获取。

评论图由不同类型的节点组成，如图 3 所示。

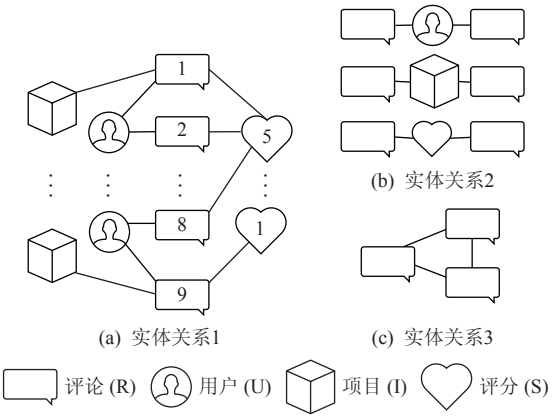


Fig. 3 Multi-view partitioning explanation

图 3 多视图划分说明

本文利用用户、商品和评分的结构信息对评论信息进行针对性建模，基于 3 种元路径（R—U—R, R—I—R, R—S—R），将原始异构网络转化为只包含评论信息的 3 个网络视图（以用户为关系的评论视图，以商品为关系的评论视图，以评分为关系的评论视图）。

基于以上论述，多视图划分过程伪代码如算法 1 所示。

算法 1. 多视图划分实现过程。

输入：四元组列表 quadruples (review *ID*, user *ID*, item *ID*, *score*);

输出：3 个评论视图。

① 初始化 $\boldsymbol{A}_1 = []$; $\boldsymbol{A}_2 = []$; $\boldsymbol{A}_3 = []$;

② for *i* in *range*(*quadruples.count*()) do

③ review ID_1, user ID_1, item ID_1, $score_1$ = *quadruples*[*i*];

④ for *j* in *range*(*quadruples.count*())[*i*:] do

⑤ reivew ID_2, user ID_2, item ID_2, $score_2$ = *quadruples*[*j*];

⑥ if user ID_1 == user ID_2 && (review ID_1, review ID_2) not in $\boldsymbol{A}_1$

⑦ $\boldsymbol{A}_1$.*append*(review ID_1, review ID_2);

⑧ end if

⑨ if item ID_1 == item ID_2 && (review ID_1, review ID_2) not in $\boldsymbol{A}_2$

⑩ $\boldsymbol{A}_2$.*append*(reivew ID_1, reivew ID_2);

⑪ end if

⑫ if $score_1$ == $score_2$ && (review ID_1, review ID_2) not in $\boldsymbol{A}_3$

⑬ $\boldsymbol{A}_3$.*append*(review ID_1, review ID_2);

⑭ end if

⑮ end for

⑯ end for

⑰ return $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, $\boldsymbol{A}_3$。

这种转化过程可以帮助网络聚焦于评论信息，通过分析评论与用户、项目和评分之间的关系，使网络拥有对于评论节点更深入的理解和洞察力。

将每个评论的属性信息X_{r}、发布此评论的用户属性信息X_{u}、评论对应的项目属性信息X_{i}以及评论的评分属性信息X_{s}进行拼接，得到融合多种信息的节点特征X_{v_i}，如式（1）所示。

$$X_{v_i} = [X_{\mathrm{r}} \| X_{\mathrm{u}} \| X_{\mathrm{i}} \| X_{\mathrm{s}}] \text{。} \tag{1}$$

使用 MLP 层将融合过多信息的特征表示映射到高维空间中，得到的特征表示$\boldsymbol{H}^{(0)}$作为评论节点的初始表示，具体计算过程如式（2）所示。

$$\boldsymbol{H}^{(0)} = \boldsymbol{MLP}(X) \text{。} \tag{2}$$

2.2.2 多视图邻域差异聚合模块

该模块合并邻域信息与自身-邻域差异信息，使网络同时关注节点之间的关联性与差异性，提高了虚假评论节点评分器对于虚假评论的敏感性。

在上文中，使用多视图划分与多信息融合模块得到多个视图以及节点特征$\boldsymbol{H}^{(0)}$。不同视图中节点链接的邻居各不相同，因此首先需要进行单视图邻域聚合。邻域特征使用传统图卷积网络的方式进行聚合，而自身-邻域差异特征通过自身特征与图卷积网络得到的邻域特征相减得到。再将邻域差异特征与邻域特征进行拼接，从而进一步强调评论与评论之间的关联与差异，如式（3）所示。

$$\boldsymbol{H}_{i,r}^{(l)} = \left(\boldsymbol{AGGR}\{\boldsymbol{H}_j^{(l)}, j \in N_r(i)\} \| \boldsymbol{H}_i^{(l)} - \boldsymbol{AGGR}\{H_j^{(l)}, j \in N_r(i)\}\right), \tag{3}$$

其中$\boldsymbol{H}_i^{(l)}$表示第l层中第i个节点自身信息，$\boldsymbol{H}_j^{(l)}$表示第i个节点的邻域节点的信息，$v_j \in N_r(i)$即v_j是节点v_i在关系r下的邻居。$\boldsymbol{AGGR}\{\cdot\}$的具体计算过程如式（4）所示。

$$\boldsymbol{AGGR}\left\{\boldsymbol{H}_j^{(l)}, j \in N_r(i)\right\} = \frac{\sum_{j=1}^{N_r} \boldsymbol{H}_j^{(l)}}{|N_r|} \text{。} \tag{4}$$

得到单视图聚合邻居信息的节点表示后,为了获得节点更全面的信息,需要融合不同视角的信息,以获得高阶的语义信息。每个视图的注意力权重计算如式(5)所示,利用 softmax 函数对所有视图的重要程度进行归一化,其中r表示不同的视图,$\boldsymbol{\varphi}_r$是可学习共享参数矩阵,$\alpha_r^{(l)}$是注意力系数。

$$\alpha_r^{(l)}=\frac{\exp\left(\boldsymbol{H}_{i,r}^{(l)}\boldsymbol{\varphi}_r\right)}{\sum_{r=1}^{R}\exp\left(\boldsymbol{H}_{i,r}^{(l)}\boldsymbol{\varphi}_r\right)}。\tag{5}$$

将学习到的注意力系数作为权重,多视图的节点嵌入融合如式(6)所示。

$$\boldsymbol{H}_i^{(l+1)}=\sum_{r=1}^{R}a_r^{(l)}\boldsymbol{H}_{i,r}^{(l)}。\tag{6}$$

根据对图卷积运算的研究[26],不同神经网络层输出的节点表示具有不同的平滑度。为减少平滑度对于网络的影响,结合不同神经网络层输出的中间表示可以进一步提高虚假评论检测的质量,节点最终表示如式(7)所示。

$$\boldsymbol{H}_i^{(\text{final})}=\left[\boldsymbol{H}_i^{(0)}\|\boldsymbol{H}_i^{(1)}\|\cdots\|\boldsymbol{H}_i^{(L)}\right],\tag{7}$$

其中$\boldsymbol{H}_i^{(\text{final})}\in\mathbb{R}^{(2^{L+1}-1)d}$。

2.2.3 虚假评论评分器及偏差损失

最终的节点表示$\boldsymbol{H}^{(\text{final})}$将被输入到虚假评论节点评分器中,进一步评估每个评论节点的异常程度。虚假评论节点评分器由 2 个全连接层构成,将中间节点表示转换为标量异常分数,如式(8)所示。

$$o_i=ReLU\left(\boldsymbol{W}_s\boldsymbol{H}_i^{(\text{final})}+\boldsymbol{b}_{s_1}\right),\ s_i=\boldsymbol{u}_s^{\mathrm{T}}o_i+\boldsymbol{b}_{s_2},\tag{8}$$

其中s_i是节点l的异常分数,o_i是中间输出。$\boldsymbol{W}_s$和$\boldsymbol{u}_s$分别是可学习的权重矩阵和权重向量。$\boldsymbol{b}_{s_1}$和$\boldsymbol{b}_{s_2}$是相应的偏置项。

鉴于参考偏差损失在过去图异常检测研究中的成功应用[27],本文采用偏差损失作为损失函数,以引导模型在异常分数集合中统计分布的层面区分正常和异常评论。

高斯分布通常可以拟合数据集中正常节点的分数[28],为了指导模型学习,从高斯先验分布中随机采样 k 个分数,即参考分数$SR=(sr_1,sr_2,\cdots,sr_k)\sim N(\mu,\sigma^2)$,作为正常评论节点在统计分布上的先验参考分数(均值μ_{sr}和标准差σ_{sr})计算如式(9)(10)所示。

$$\mu_{sr}=\frac{1}{k}\sum_{i=1}^{k}sr_i,\tag{9}$$

$$\sigma_{sr}=\sqrt{\frac{\sum_{i=1}^{k}(sr_i-\mu_{sr})^2}{k-1}}。\tag{10}$$

利用参考分数SR,用标准分数的形式定义节点v_i的分数和参考分数 SR 之间的偏差dev,具体如式(11)所示。

$$dev(v_i)=\frac{s_i-\mu_{sr}}{\sigma_{sr}}。\tag{11}$$

最终的偏差损失函数 L 如式(12)所示。

$$L=(1-y_i)|dev(v_i)|+y_i\max(0,m-|dev(v_i)|),\tag{12}$$

其中y_i是输入节点v_i的真实标签,如果v_i为虚假评论节点,则y_i=1,否则y_i=0;m是一个置信边界,定义了偏差的半径。

通过最小化上述损失函数,将使正常节点的异常分数尽可能接近μ_{sr},同时使μ_{sr}与异常节点的异常分数之间产生m的偏差。因此,如果一个节点的模式明显偏离正常模式,那么它将被赋予一个大的异常分数。

此外,由于无法访问正常节点的标签,此处简单地将V^{μ}中未标记的节点视为正常节点。通过这种方式,剩余的未标记虚假节点和所有正常节点都被视为正常,从而将污染引入训练集。3.7 节评估了不同污染水平对模型性能的影响,实验表明本文模型对不同水平的污染具有鲁棒性。

2.2.4 元学习

元学习利用多个辅助网络模拟小样本问题对网络进行训练,提取不同网络中的元知识,再用目标网络进行进一步的微调,使得目标网络在小样本场景下表现较好,泛化性较高。

本文将每个辅助学习任务定义为在单个网络上执行小样本虚假评论检测,利用元学习中的 support 集模拟目标网络中的具体场景。设T_i表示由网络G_i^{S}构建的小样本虚假评论检测任务,则每个 epoch 有 P 个辅助学习任务,f_θ表示参数为θ的 MGDN 模型,在每个 epoch 初始时,所有辅助网络和目标网络的参数均为θ。在给定 P 个辅助任务的情况下,每个辅助任务划分为 support 集和 query 集,support 集用于辅助网络自身更新,query 集用于验证辅助网络并更新目标网络。辅助网络通过对 support 集梯度更新来更新参数,此过程如式(13)所示。

$$\theta_i'=\theta-\alpha\nabla_\theta L_{T_i}(f_\theta),\tag{13}$$

其中α为元学习内学习率,L_{T_i}为式(12)中的损失函数。

式(13)只包括 1 步梯度更新,可以扩展为多步梯度更新,使用 query 集对于更新后的网络进行测试,并将所有辅助网络测试得到的损失整合,如式(14)所示。

$$\frac{1}{P}\sum_{i=1}^{P}L_{T_i}\left(f_{\theta_i'}\right)=\frac{1}{P}\sum_{i=1}^{P}L_{T_i}\left(f_{\theta-\alpha\nabla_\theta L_{T_i}(f_\theta)}\right)。\tag{14}$$

根据式(14)得到的损失对 MGDN 的目标网络进行反向更新，更新的模型参数能够学习到每个辅助网络在虚假评论检测中的能力，提取辅助网络所提供的共性知识。这样，目标网络可以快速适应和学习针对目标任务的特征。

具体而言，使用随机梯度下降(SGD)来更新目标网络中的模型参数θ，其中β为元学习外学习率，如式(15)所示。

$$\theta\leftarrow\theta-\beta\nabla_\theta\frac{1}{P}\sum_{i=1}^{P}L_{T_i}\left(f_{\theta_i'}\right)。\tag{15}$$

最后，通过使用目标网络的训练集对式(15)中辅助网络训练后更新的网络参数进行微调，以进一步提高其适应目标任务的能力。

3 实验与分析

3.1 实验数据集

本文在公开数据集 Yelp[29] 及 2 个来自亚马逊的真实评论数据集 Sports 和 Toys 上进行各项实验来验证 Meta-MGDN 的性能，如表 1 所示。其中，Yelp 数据集收集了 Yelp 上的酒店和餐厅评论，Sports 数据集涵盖了运动户外产品的评论信息，Toys 数据集包含了玩具与游戏产品分区的评论信息。除了评论内容本身，数据集还包括评论 id、用户 id、项目 id、评分以及标签信息。

Table 1 Statistics of Dataset

表 1 数据集统计

数据集	节点数	不平衡率	关系	关系边数
Yelp	10 530	12.48	R—U—R	34 640
			R—I—R	66 368
			R—S—R	429 540
Toys	9 824	9.60	R—U—R	3 556
			R—I—R	243 966
			R—S—R	18 230 829
Sports	9 943	8.37	R—U—R	4 629
			R—I—R	3 211 575
			R—S—R	18 702 979

注：R 表示评论，U 表示用户，I 表示项目，S 表示评分。

本文参考文献 [30] 的设置方法，采用 LDA(latent Dirichlet allocation)主题模型来提取评论节点初始属性，并将 LDA 主题个数设置为 32。批处理大小均设置为 200，网络层数设置为 3，辅助网络个数设置为 3，内学习率设置为 0.001，外学习率设置为 0.000 3。

3.2 实验对比方法

为了评估模型的虚假评论检测性能，本文选取了基线方法以及几种先进方法进行对比实验。

1)DevNet[27]。该方法通过偏差学习实现异常分数的端到端学习，利用少量标记异常和先验概率使得异常分数与正常数据的统计显著偏差以进行异常检测。

2)SemiGNN[31]。该方法是一种半监督的 GNN 模型，利用层次化注意力机制更好地关联不同的邻居和不同的视图进行节点表示，从而进行异常检测。

3)DOMINANT[18]。该方法是一种基于图卷积网络的自编码器框架，通过网络结构和节点属性的重构误差来进行异常检测。

4)Meta-GDN[21]。该方法是一种基于元学习的 GNN 模型，能够从目标网络所在领域的多个辅助网络中提取元知识，进行小样本的异常检测。

5)GraphCAD[3]。该方法利用图对比学习，通过将异常节点与正常节点在其与全局上下文的距离方面进行对比，进而判断节点的异常性。

6)GTAN[32]。该方法是一种半监督欺诈检测的图神经网络模型，通过构建事务图并利用门控注意力网络来学习事务表示，进行异常检测。

3.3 评价指标

本文为全面评估不同虚假评论检测方法的性能，使用在先前的研究中被广泛认可的评价指标[27-31]：

1)AUC-ROC。代表了随机选取的一个异常样本相较于随机选取的一个正常样本，模型预测得到更高分数的概率。

2)AUC-PR。表示 PR(precision-recall)曲线下的面积，反映模型在识别异常样本时的整体表现。

3)Precision@K。反映在前 K 个预测为异常的样本中，真实样本和异常样本所占的比例。

3.4 对比实验

Meta-MGDN 与基线方法在本文选用的数据集上的对比实验结果如表 2 所示。经过实验验证，本文通过与基准方法比较，评估了 Meta-MGDN 的性能。在表 2 中，展示了关于 AUC-ROC 和 AUC-PR 的评估结果(10 个样本)，由实验结果可看出，在给定评价指标上，Meta-MGDN 取得了最好的检测效果。同时在图 4 中可视化了关于 Precision@K 的结果。

根据这些结果，得出以下结论：1)就 AUC-ROC 和 AUC-PR 而言，Meta-MGDN 方法相较于其他方法

Table 2 Comparative Experimental Results for 10 Samples

表 2 10 个样本的对比实验结果

方法	Yelp 数据集		Toys 数据集		Sports 数据集	
	AUC-ROC	AUC-PR	AUC-ROC	AUC-PR	AUC-ROC	AUC-PR
DevNet	0.573	0.225	0.515	0.212	0.567	0.206
SemiGNN	0.531	0.105	0.552	0.153	0.561	0.163
DOMINANT	0.489	0.155	0.513	0.213	0.521	0.225
Meta-GDN	0.781	0.583	0.802	0.602	0.813	0.621
GraphCAD	0.713	0.405	0.706	0.413	0.721	0.451
GTAN	0.793	0.561	0.812	0.613	0.821	0.632
Meta-MGDN（本文）	0.815	0.616	0.857	0.656	0.865	0.737

展现出显著的优势。与最新的方法相比，在不同数据集上，Meta-MGDN 方法在 AUC-ROC 上提升了 2.2%~4.4%，在 AUC-PR 上则提升了 4.3%~10.5%。此外，Precision@K 的结果再次验证了 Meta-MGDN 可以准确地评估异常分数，能够更好地对异常节点进行排序。2)无监督方法 DOMINANT 性能受限于无法利用标记异常的监督知识。在不同数据集上，Meta-MGDN 在 AUC-ROC 指标上提升 32.6~34.4 个百分点，AUC-PR 指标提升 44.3~51.2 个百分点。3)有监督、半监督方法如 DevNet, SemiGNN, GTAN 等也未能提供令人满意的结果。这些模型在初始属性表示时，将不同类别的节点映射到同一空间，会导致信息缺失的问题，且在同类型节点聚合时，未能捕捉节点的个体差异性。除了 Meta-GDN 方法外，其他方法均缺乏应对样本极度稀缺的策略，都依赖于大量的标记数据，这限制了它们在小样本环境下的性能。而 Meta-GDN 虽然解决了样本稀缺问题，但由于存在信息丢失和个体差异性捕捉不足的问题，所以在节点表示能力上仍有局限。因此，本文提出的 Meta-MGDN 在性能上表现更加优越。

3.5 消融实验

为验证 Meta-MGDN 的元学习框架、信息融合模块以及多视图邻域差异聚合模块的有效性，本文进行了消融实验，实验结果如图 5(a)(b)所示。其中，Meta-MGDN 表示完整模型；w/o Meta 表示不使用元学习架构，相较于原方法，AUC-ROC 和 AUC-PR 指标在 3 个数据集上分别下降了 5~5.6 个百分点和 18.3~24.3 个百分点；w/o Fuse 表示评论节点不融合多重信息，只使用评论节点本身信息，相较于 Meta-MGDN，AUC-ROC 和 AUC-PR 指标在 3 个数据集上分别下降了 3.2~20.4 个百分点和 14.8~45.5 个百分点；w/o Differ 表示邻域聚合时不使用差异性聚合，而使用传

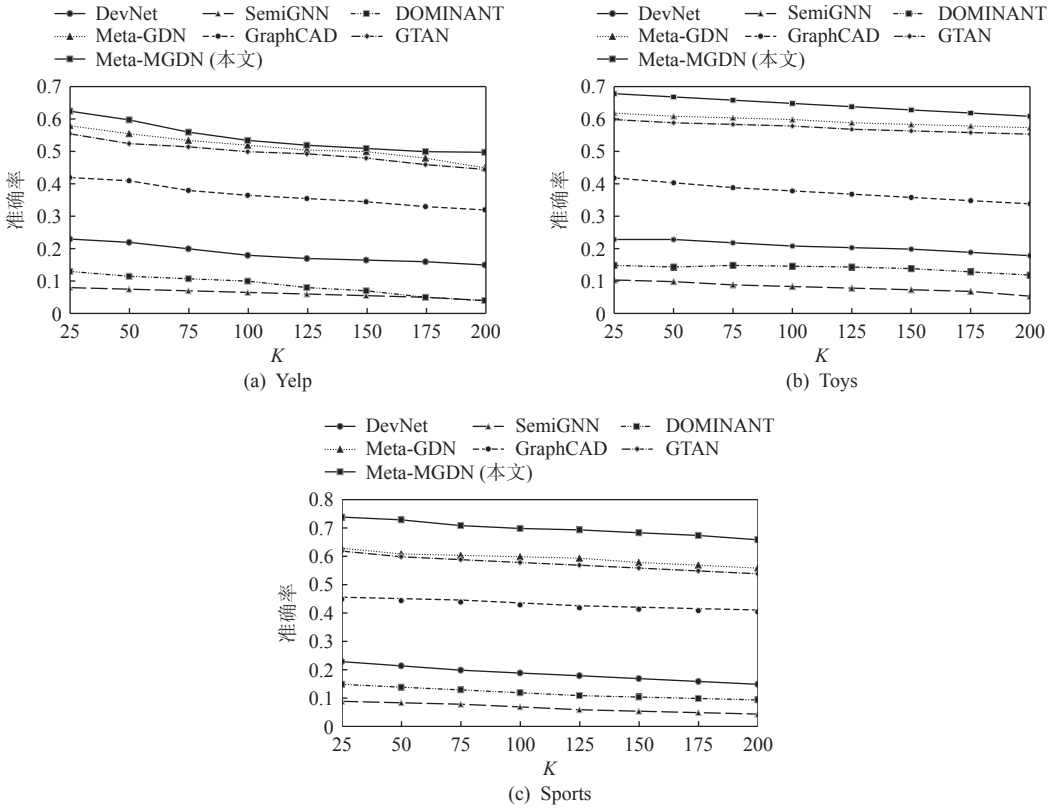
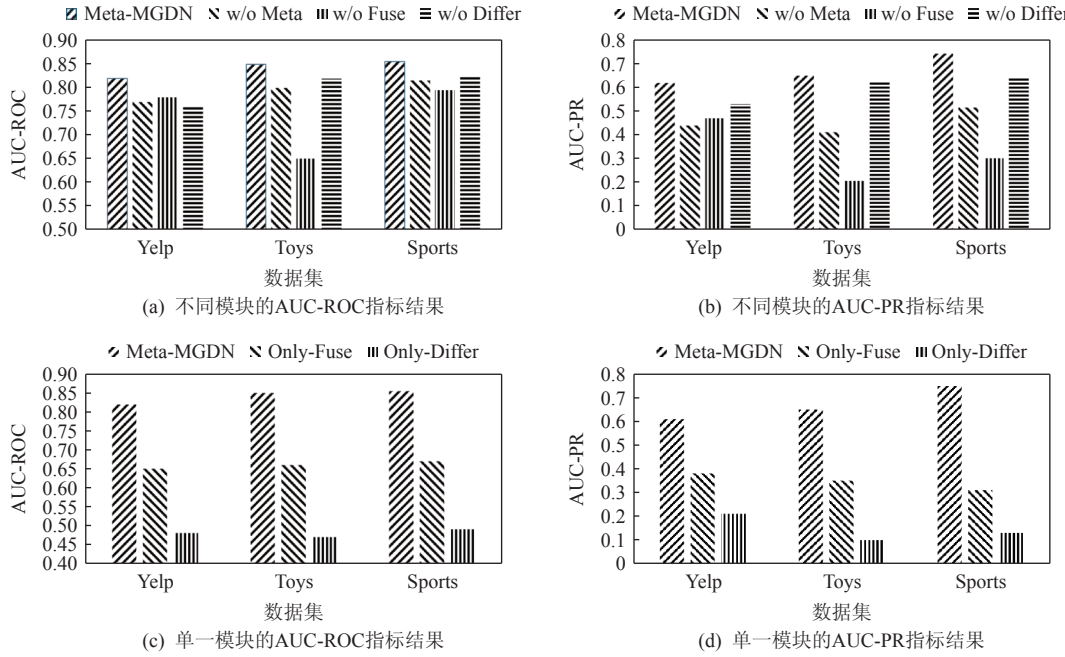


Fig. 4 Comparative results of Precision@K on ten samples in three datasets

图 4 3 个数据集上在 10 个样本的 Precision@K 的对比结果



(a) 不同模块的AUC-ROC指标结果　(b) 不同模块的AUC-PR指标结果

(c) 单一模块的AUC-ROC指标结果　(d) 单一模块的AUC-PR指标结果

Fig. 5　Results of ablation study metrics

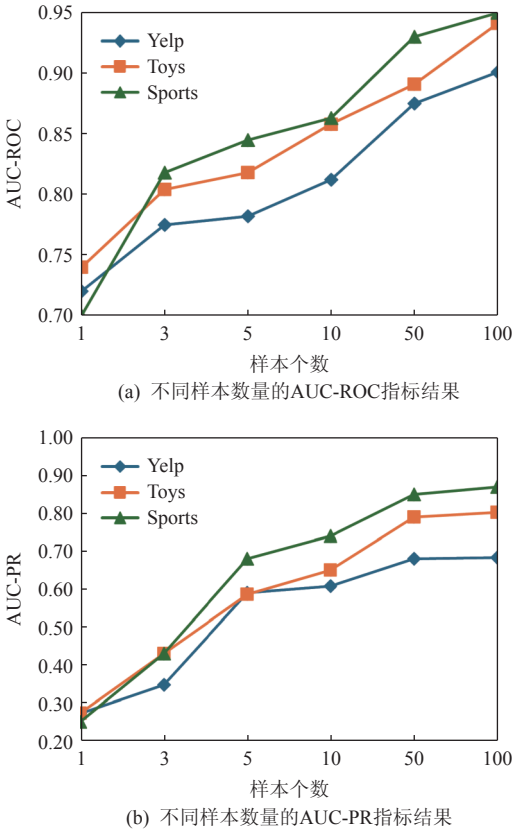
图 5　消融实验指标结果

统图卷积网络的邻域聚合，相较于 Meta-MGDN，AUC-ROC 和 AUC-PR 指标在 3 个数据集上分别下降了 2.4~4.7 个百分点和 3.5~10.2 个百分点。

为验证 Meta-MGDN 中单一模块的有效性，本文设计了 2 个模型变体进行消融实验，实验结果如图 5(c)(d)所示。其中，Only-Fuse 表示模型去除元学习和差异聚合模块，仅使用多重信息融合模块和图卷积网络的邻域聚合；Only-Differ 表示模型去除元学习和多重信息融合模块，仅使用多视图邻域差异聚合模块。实验结果表明，相较于 Meta-MGDN，2 个变体的指标均有大幅下降，并且都低于图 5(a)(b)的 3 个变体，侧面验证了单一模块的有效性。

3.6　样本数量控制实验

为验证 Meta-MGDN 在不同样本个数的虚假评论检测任务中的有效性，本文评估了 Meta-MGDN 在目标网络上针对不同标记异常样本数量的性能(单样本、3 个样本、5 个样本、10 个样本、50 个样本和 100 个样本)，实验结果如图 6 所示。通过与表 2 的结果进行比较，可以观察到即使只有 1 个标记异常样本(即单样本)时，Meta-MGDN 仍然展现出了良好的性能，AUC-ROC 指标分别在 3 个数据集上可以达到 0.72, 0.73, 0.70，并且明显优于大多数 10 个标记异常样本的基线方法。同时，可以清晰地看到随着标记异常样本数量的增加，Meta-MGDN 的性能也得到了不同程度的提升，这表明通过更多的标记样本微调目标网络，可以进一步改善 Meta-MGDN 的性能表现，且在 100 个样本的时候，模型的 AUC-ROC 和 AUC-



(a) 不同样本数量的AUC-ROC指标结果

(b) 不同样本数量的AUC-PR指标结果

Fig. 6　Results of sample control experiment metrics on three datasets

图 6　3 个数据集上样本控制实验指标结果

PR 指标可以达到 0.95 和 0.86。

3.7 鲁棒性实验

正如 2.2.3 节所讨论的，模型将从未标记的数据中采样的所有节点都视为正常节点用于计算偏差损失。这种简单的策略在未标记的训练数据中引入了异常污染。

本文对比了 Meta-MGDN 和基线方法对不同污染程度(即未标记训练数据中的异常比例)的鲁棒性，实验结果如图 7 所示。经实验得知，Meta-MGDN 方法在污染率 0~0.2 的情况下仅损失了 2.7% 的 AUC-ROC 指标，而其他方法在此实验场景下分别下降了 5.2~21.8 个百分点。尽管所有方法的性能随着污染程度的增加而下降，但 Meta-MGDN 在很大程度上都表现出了显著的鲁棒性，并且始终明显优于其他方法。

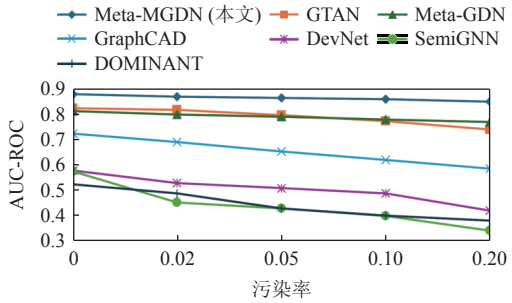


Fig. 7 Results of AUC-ROC metrics for different levels of pollution

图 7 不同污染程度的 AUC-ROC 指标结果

3.8 实验环境设置与时间评估

本文的实验代码使用 PyTorch 深度学习框架开发，在 Ubuntu 操作系统上搭建实验环境，具体实验环境配置信息如表 3 所示。

Table 3 Experimental Environment Configuration

表 3 实验环境配置

配置名称	配置参数
中央处理器	AMD Ryzen™ 5 2600 CPU @ 3.4GHz
内存/GB	128
图形处理器	NVIDIA GeForce GTX 1080TI
操作系统	Ubuntu 18.04.2 LTS
Pytorch 版本	1.10.1 + cu11.3

为了评估 Meta-MGDN 模型的性能提升是否以较大的时间复杂度为代价，本文采用相同规模的训练集，将 Meta-MGDN 与其他基线方法在训练周期内的平均消耗时间进行了对比，并将每个模型对应的 AUC-ROC 值作为衡量模型性能的重要指标，实验结果图 8 所示。

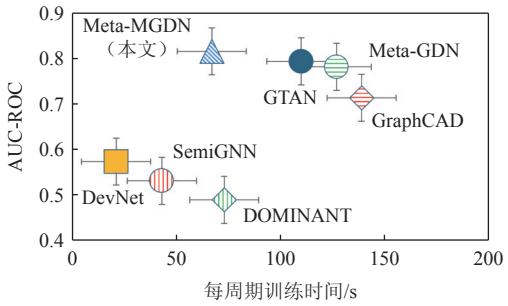


Fig. 8 Comparison of performance and training time among different models

图 8 不同模型性能及训练时间对比

分析图 8 的实验结果可以看出 Meta-MGDN 在模型性能和训练速度方面都表现出明显的优势。首先，从时间消耗的角度，Meta-MGDN 相对于其他模型在训练周期内的平均消耗时间较低。验证 Meta-MGDN 能够在较短的时间内完成训练过程。其次，通过对比 AUC-ROC 指标，可以看出 Meta-MGDN 在模型性能上也取得了显著提升。因此，Meta-MGDN 能够在保持较低时间复杂度的同时提升虚假评论检测性能。

4 结 论

针对虚假评论检测任务中普遍存在的样本标记稀缺问题，本文提出一种基于元学习框架的解决方案，旨在有效缓解因样本不足导致的训练困难问题，增强模型在小样本场景下的泛化能力。通过在公开评论数据集上进行的实验验证了本文所提模型在虚假评论检测任务上的优越性能。未来研究工作将聚焦于进一步完善语言特征的识别，探索如何更全面地整合评论相关的异质信息，如时序信息等，进一步提升虚假评论检测模型的性能与准确性。

参 考 文 献

[1] Ott M, Cardie C, Hancock J. Estimating the prevalence of deception in online review communities[C]//Proc of the 21st Int Conf on World Wide Web. New York: ACM, 2012: 201−210

[2] López V, Río S D, Benítez J M, et al. Cost-sensitive linguistic fuzzy rule based classification systems under the MapReduce framework for imbalanced big data[J]. Fuzzy Sets & Systems, 2015, 258: 5−38

[3] Chen Bo, Zhang Jing, Zhang Xiaokang, et al. GCCAD: Graph contrastive coding for anomaly detection[J]. IEEE Transactions on Knowledge & Data Engineering, 2022, 35(8): 8037−8051

[4] Chai Ziwei, Yang Yang, Dan Jiawang, et al. Towards learning to

discover money laundering sub-network in massive transaction network[C]//Proc of the 35th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2023: 14153−14160

[5] Liu Jiawei, Shi Chuan, Yang Cheng, et al. A survey on heterogeneous information network based recommender systems: Concepts, methods, applications and resources[J]. AI Open, 2022, 3: 40−57

[6] Mohawesh R, Xu Shuxiang, Tran S, et al. Fake reviews detection: A survey[J]. IEEE Access, 2021, 9: 65771−65802

[7] Shen Xiao, Dai Quanyu, Chung F L, et al. Adversarial deep network embedding for cross-network node classification[J]. arXiv preprint, arXiv: 2002.07366, 2020

[8] Huang Mengda, Liu Yang, Ao Xiang, et al. AUC-oriented graph neural network for fraud detection[C]//Proc of the 31st ACM Web Conf. New York: ACM, 2022: 1311−1321

[9] Alhosseini S A, Tareaf R B, Najafi P, et al. Detect me if you can: Spam bot detection using inductive representation learning[C]//Proc of the 28th World Wide Web Conf. New York: ACM, 2019: 148−153

[10] Yang Ruichao, Wang Xiting, Jin Yiqiao, et al. Reinforcement subgraph reasoning for fake news detection[C]//Proc of the 28th ACM SIGKDD Conf on Knowledge Discovery and Data Mining. New York: ACM, 2022: 2253−2262

[11] Ghadery E, Movahedi S, Faili H, et al. MNCN: A multilingual ngram-based convolutional network for aspect category detection in online reviews[C]//Proc of the 33rd AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2019: 6441−6448

[12] Wu Man, Pan Shirui, Zhou Chuan, et al. Unsupervised domain adaptive graph convolutional networks[C]//Proc of the 31st Web Conf. New York: ACM, 2020: 1457−1467

[13] Yang Jiajie, Mao Dianhui, Zhang Denghui. Graph attention network-based multimodal fusion for fake food reviews detection[C]//Proc of the 9th Int Conf on Cyber Security and Information Engineering. New York: ACM, 2024: 233−241

[14] Huang Kexin, Zitni M. Graph meta learning via local subgraphs[C]//Proc of the 33rd Conf on Neural Information Processing Systems. New York: ACM, 2020: 5862−5874

[15] Yao Huaxiu, Zhang Chuxu, Wei Ying, et al. Graph few-shot learning via knowledge transfer[J]. arXiv preprint, arXiv: 1910.03053, 2020

[16] Ding Kaize, Zhou Qinghai, Tong Hanghang, et al. Few-shot network anomaly detection via cross-network meta-learning[C]//Proc of the 30th Int Conf of World Wide Web. New York: ACM, 2021: 2448−2556

[17] Cao Shaosheng, Lu Wei, Xu Qiongkai. Deep neural networks for learning graph representations[C]//Proc of the 13th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2016: 1145−1152

[18] Ding Kaize, Li Jundong, Bhanushali R, et al. Deep anomaly detection on attributed networks[C]//Proc of the 2019 SIAM Int Conf on Data Mining. Philadelphia, PA: SIAM, 2019: 594−602

[19] Zhou Fan, Cao Chengtai, Zhang Kunpeng, et al. Meta-GNN: On few-shot node classification in graph meta-learning[C]//Proc of the 28th ACM Int Conf on Information and Knowledge Management. New York: ACM, 2019: 2357–2360

[20] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[J]. arXiv preprint, arXiv: 1609.02907, 2016

[21] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks[J]. arXiv preprint, arXiv: 1710.10903, 2017

[22] Hamilton W, Ying Zhitao, Leskovec J. Inductive representation learning on large graphs[C]//Proc of the 31st Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2017: 1025−1035

[23] He Junwei, Xu Qianqian, Jiang Yangbangyan, et al. ADA-GAD: Anomaly-denoised autoencoders for graph anomaly detection[C]// Proc of the 38th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2024: 8481−8489

[24] Roy A, Shu Juan, Li Jia, et al. GAD-NR: Graph anomaly detection via neighborhood reconstruction[C]//Proc of the 17th ACM Int Conf on Web Search and Data Mining. New York: ACM, 2024: 576−585

[25] Cai Jinyu, Zhang Yunhe, Lu Zhoumin, et al. Towards effective federated graph anomaly detection via self-boosted knowledge distillation[C]//Proc of the 32nd ACM Int Conf on Multimedia. New York: ACM, 2024: 5537−5546

[26] Wu F, Souza A, Zhang Tianyi, et al. Simplifying graph convolutional networks[J]. arXiv preprint, arXiv: 1902.07153, 2019

[27] Pang Guansong, Shen Chunhua, Hengel A, et al. Deep anomaly detection with deviation networks[C]//Proc of the 25th ACM SIGKDD Int Conf on Knowledge Discovery & Data Mining. New York: ACM, 2019: 353−362

[28] Kriegel H P, Kröger P, Schubert E, et al. Interpreting and unifying outlier scores[C]//Proc of the 2011 SIAM Int Conf on Data Mining. Philadelphia, PA: SIAM, 2011: 13−24

[29] Rayana S, Akoglu L. Collective opinion spam detection: Bridging review networks and metadata[C]//Proc of the 21st ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2015: 985−994

[30] Wang Ye, Nagireddy S R, Thota C T, et al. Community-in-the-loop: Creating artificial process intelligence for co-production of city service[C]//Proc of the ACM on Human-Computer Interaction (CSCW2). New York: ACM, 2022: 1−21

[31] Wang Daixin, Lin Jianbin, Cui Peng, et al. A semi-supervised graph attentive network for financial fraud detection[C]//Proc of the 2019 IEEE Int Conf on Data Mining. Piscataway, NJ: IEEE, 2019: 598−607

[32] Xiang Sheng, Zhu Mingzhi, Cheng Dawei, et al. Semi-supervised credit card fraud detection via attribute-driven graph representation[C]//Proc of the 37th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2023: 14557−14565



Chen Zhuo, born in 1978. PhD, associate professor. Her main research interests include natural language processing and anomaly detection. (chenzhuo_qust@163.com)

陈　卓, 1978 年生。博士，副教授。主要研究方向为自然语言处理、异常检测。



Di Jiaqi, born in 1999. Master. His main research interests include natural language processing and anomaly detection.

狄家奇, 1999 年生。硕士。主要研究方向为自然语言处理、异常检测。



Zhang Ao, born in 2000. Master. His main research interests include recommendation algorithms and anomaly detection.

张 傲, 2000 年生。硕士。主要研究方向为推荐算法、异常检测。



Lu Xiaopeng, born in 2001. Master candidate. His main research interest includes natural language processing.

卢晓鹏, 2001 年生。硕士研究生。主要研究方向为自然语言处理。



Du Junwei, born in 1974. PhD, professor. His main research interests include intelligent software engineering, intelligent safety in chemical engineering, graph neural networks, and recommender systems. (djwqd@163.com)

杜军威, 1974 年生。博士，教授。主要研究方向为智能软件工程、智能化工安全、图神经网络、推荐系统。