文章编号：1003-0077(2023)06-0137-10

基于对抗图增强对比学习的虚假新闻检测

陈卓敏[1]，王 莉[1]，朱小飞[2]，王子康[3]

(1. 太原理工大学 计算机科学与技术学院(大数据学院)，山西 晋中 030600；
2. 重庆理工大学 计算机科学与工程学院，重庆 400054；
3. 太原理工大学 软件学院，山西 晋中 030600)

摘 要：随着互联网的快速发展，社交媒体成为了新闻发布和传播的主要平台，如何准确识别虚假新闻已成为研究热点。现有的基于深度学习的虚假新闻检测方法在面对噪声和敌对信息时缺乏鲁棒性。为了应对这一挑战，该文提出了一种对抗图增强对比学习的方法，该方法引入对抗对比学习，使模型抓住少量但充分的信息完成增强图与原始图之间的互信息最大化，在进行训练时重点捕捉有用信息。同时，该模型还利用了特征增强器和图表示对比学习进行图表示增强，加强特征学习。在两个公共数据集上进行的实验表明，该模型在现有基线上达到了最优的性能。

关键词：虚假新闻检测；对比学习；对抗图增强；社交网络

中图分类号：TP391 **文献标识码**：A

Fake News Detection Based on Adversarial Graph Enhanced Contrastive Learning

CHEN Zhuomin[1], WANG Li[1], ZHU Xiaofei[2], WANG Zikang[3]

(1. College of Computer Science and Technology & College of Data Science, Taiyuan University of Technology, Jinzhong, Shanxi 030600, China;
2. College of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400054, China;
3. College of Software, Taiyuan University of Technology, Jinzhong, Shanxi 030600, China)

Abstract: Existing deep learning-based fake news detection methods are defected in dealing with face of noise and hostile information. This paper proposes an adversarial graph-enhanced contrastive learning method to address this issue. It introduces adversarial contrastive learning so that the model can maximize the mutual information between the enhanced graph and the original graph. Specifically, this model also applies a feature enhancer to the graph representation contrastive learning. Experimental results on two public datasets show that the proposed model achieves state-of-the-art performance.

Keywords: fake news detection; contrastive learning; adversarial graph enhancement; social networks

0 引言

社交媒体使虚假新闻的传播速度加快，扰乱了网络秩序，误导了公众，影响了正常的个人生活和社会秩序[1-2]。因此，构建有效的虚假新闻检测方法非常重要。

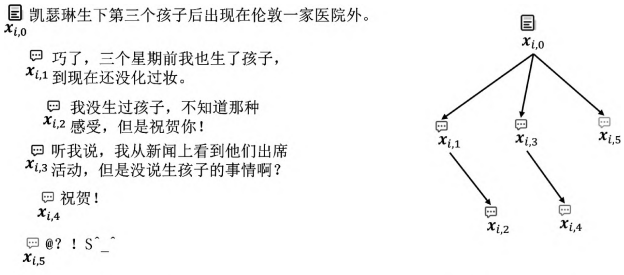
传统的虚假新闻检测方法利用有监督的分类器，如决策树[3]、随机森林[4]和支持向量机(SVM)[5]。虽然这些方法可以在一定程度上提高虚假新闻检测的准确度，但主要依赖于费时费力的手工制作特征。最近，基于深度学习的方法(如RNN[6]、RvNN[7])从虚假新闻传播中捕获序列特征，或者通过卷积神经网络(如CNN[8]、Bi-GCN[9]、

收稿日期：2022-10-20 **定稿日期**：2022-11-21

基金项目：国家自然科学基金(U22A20167)；国家重点研究与发展计划(2021YFB3300503)；重庆市自然科学基金(CSTB2022NSCQ-MSX1672)；重庆市教育委员会科学技术研究计划重大项目(KJZD-M202201102)

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

DiffRank[10])从新闻的传播过程中获取高级表示。He等人[11]运用新闻传播结构提出了随机性事件增强策略,与原始图进行对比学习得到新闻表征。尽管深度学习方法在虚假新闻检测中取得了成功,但上述模型往往对噪声数据缺乏鲁棒性。例如,图1(a)所示,一些用户参与评论时,会出现错位的评论或乱码等,使传播图中存在不可靠的边,这样的冗余信息会导致模型错误分类。



(a) 回复关系

(b) 新闻传播结构图

图1 新闻传播示例

为克服现有工作存在的问题,本文提出了对抗图增强对比学习模型(AGECL),其不仅利用了图数据增强和图表示增强策略丰富了新闻的表征,而且通过对抗对比学习使模型重点捕获图中有用信息。具体来说,首先通过图数据增强策略得到新闻增强图,与新闻原始传播图分别输入到GNN中,输出新闻原始图与增强图的表征后进行对抗对比学习,这种对比学习方法不仅提高了模型的对抗能力,还使得模型关注到原始传播图中的重要信息。最后,将新闻表征与特征增强器得到的表征进行融合,利用图表示对比学习以及新闻的真实标签对新闻检测模型进行微调,进行最终预测。本文的主要贡献如下:

(1) 本文将特征增强器得到的表征作为新闻检测输入的一部分,并引入图表示对比学习来帮助模型加强特征学习,进行有效的虚假新闻检测。

(2) 本文提出了AGECL框架来学习新闻的表示,将对抗对比学习首次运用到虚假新闻检测任务中,优化图神经网络中使用的图数据增强策略,使模型经过训练后捕获有用信息。

(3) 在两个公共数据集上进行的实验表明,AGECL优于其他方法。

此外,本文还证明了AGECL在进行早期检测时的性能更好。

1 相关工作

1.1 虚假新闻检测

虚假新闻检测是自然语言处理中的一项重要任务。由于虚假新闻的广泛传播影响了新闻舆论的公信力和社会安全,其相关研究受到越来越多的关注。早期关于虚假新闻检测的工作主要集中在从新闻中提取文本特征并应用传统的学习方法(如决策树[3],SVM[5])来进行检测。近年来,深度学习模型在虚假新闻检测任务中取得了较好的效果,其可以自动挖掘潜在的语义信息,同时克服了手工特征的缺点。Vaibhav等人[12]提出了一种用于虚假新闻检测的图神经网络模型,该模型对新闻中所有句子对之间的语义关系进行建模,进行虚假新闻检测。Yu等人[13]利用卷积神经网络(CNN)提取序列特征,形成重要特征之间的交互。

上述方法的局限性在于没有充分考虑新闻的传播结构。为了捕获传播结构特征,Ma等人[7]构建了基于自底向上和自顶向下传播树的递归神经网络进行新闻检测。受到图卷积网络(GCN)[14]的启发,Bian等人[9]提出了BiGCN以同时处理虚假新闻的传播与扩散。He等人[11]通过修改新闻传播结构来提取有用的新闻传播模式。但这些方法都将传播图中的边作为可靠的拓扑连接进行消息传递,忽略了不可靠关系引起的不确定性,可能会导致模型缺乏鲁棒性,降低检测性能。

1.2 对比学习

对比学习是无监督和自监督学习方法中最具代表性的一种方法,其核心思想是通过对比正样本和负样本来学习区别表征,鼓励编码器通过最大化输入的表征和学习到的表征之间的互信息(MI)进行训练。对比学习已经应用在许多不同的任务中,例如,SimCLR模型[15]使用对比学习来提高视觉表征的质量。LS_Score模型[16]通过无监督对比学习构建摘要评价系统,在对话生成中引入对比学习[17],模型能明确地感知到积极话语和消极话语之间的差异,增加了回答的多样性。此外,对比学习也促进了图结构表示学习的发展。为得到具有鲁棒性的图表示,不同的图数据增强方案相继被提出。例如,RDEA[11]集成了三种增强策略,并利用自监督对比学习辅助虚假新闻检测。本文受到对比学习的启

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

发，将对抗对比学习首次运用到虚假新闻检测任务中，并且充分利用特征增强器和图表示对比学习来丰富新闻表征。

2　问题定义

记 $C=\{C_1,C_2,\cdots,C_n\}$ 为所有新闻事件的集合。其中，C_i 为第 i 个新闻事件，n 为新闻事件个数。每一个新闻事件的集合表示为 $C_i=(\boldsymbol{x}_{i,0},\boldsymbol{x}_{i,1},\boldsymbol{x}_{i,2},\cdots,\boldsymbol{x}_{i,|V_i|},G_i)$，其中 $\boldsymbol{x}_{i,0}$ 是该事件中的新闻表征，新闻在传播过程中会产生评论。当 $j>0$ 时，$\boldsymbol{x}_{i,j}$ 表示第 j 个相关评论，$|V_i|$ 表示 C_i 中新闻与评论的总数。本文将新闻与评论、评论与评论之间的回复关系视为新闻传播的一种方式。$G_i=(V_i,E_i)$ 表示该新闻事件的传播结构图，V_i 是节点集合，E_i 是边的集合，新闻 $\boldsymbol{x}_{i,0}$ 为根节点。如图 1(a)和图 1(b)所示，如果评论 $\boldsymbol{x}_{i,2}$ 对评论 $\boldsymbol{x}_{i,1}$ 进行回复，会有一条有向边 $\boldsymbol{x}_{i,1}\rightarrow\boldsymbol{x}_{i,2}$；如果 $\boldsymbol{x}_{i,1}$ 直接对新闻 $\boldsymbol{x}_{i,0}$ 进行评论，会有一条有向边 $\boldsymbol{x}_{i,0}\rightarrow\boldsymbol{x}_{i,1}$。特征矩阵和邻接矩阵分别表示为 $\boldsymbol{X}_i\in\mathbb{R}^{|V_i|\times d}$ 和 $\boldsymbol{A}_i\in\{0,1\}^{|V_i|\times|V_i|}$。$d$ 为向量维度。

虚假新闻检测的目标是学习一个分类器 $f:C_i\rightarrow y_i$，y_i 是四个细粒度类 N、F、T、U 中的一个(即非虚假新闻、虚假新闻、真实新闻和未经验证的新闻)。

3　模型

本节将详细介绍 AGECL 模型，模型框架如图 2 所示。AGECL 模型包括三个模块：图数据增强、对抗对比学习和图表示增强。

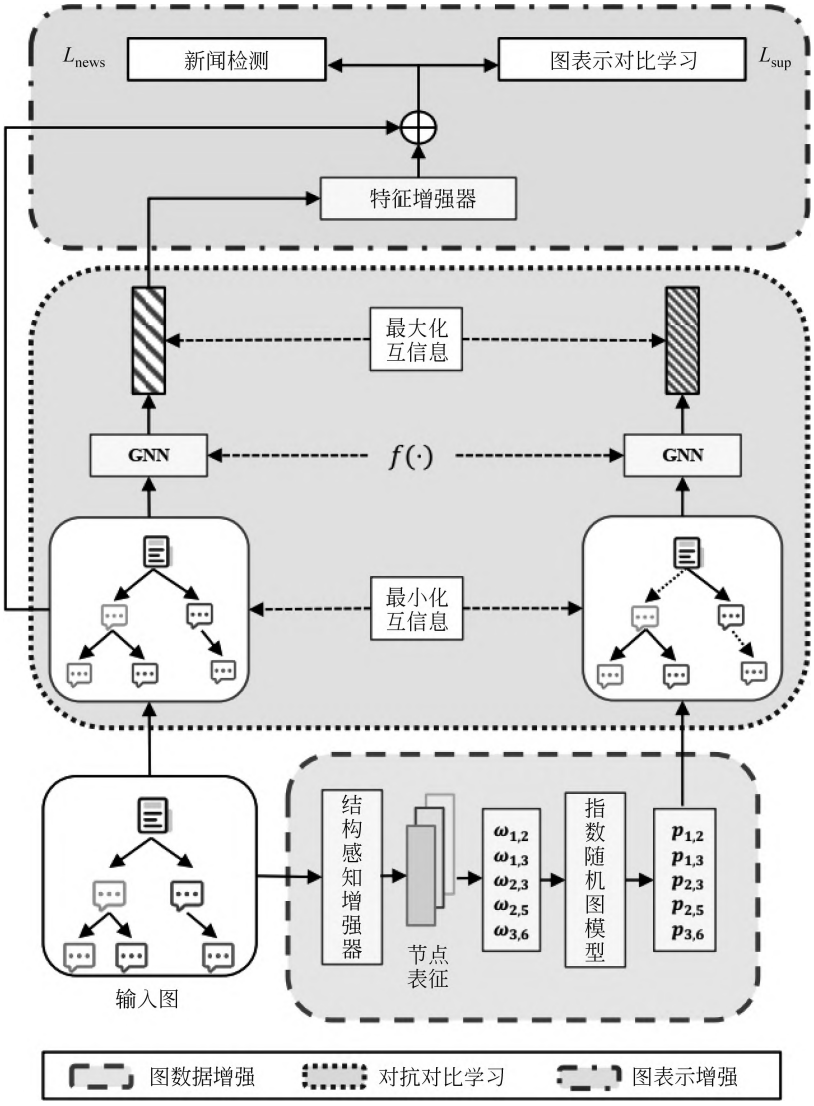


图 2　AGECL 模型框架

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

3.1 图数据增强

由于新闻的传播结构图比较庞大，会导致模型在训练时受到噪声数据的影响。因此，本文引入了边扰动方法作为图数据增强策略。边扰动即删除图中某些边。这种方法增加了除原始传播结构图以外的训练数据，减少了消息传递，同时也能减弱社交网络中的回音室效应[18]。

设图数据增强模块为 $T(\cdot)$，该模块保证进行新闻检测任务时有足够的有用信息。对于每一个新闻事件的传播图 $G=(V,E)$，记 $T(G)$，$T\in\mathcal{T}$为传播图 G 的随机图模型，$\mathcal{T}$是一个图数据增强族。每个样本 $t(G)\sim T(G)$是一个与原始传播图 G 共享同一节点集的增强图，而 $t(G)$的边集合是 E 的子集。本文将图 G 输入到结构感知增强器中，得到节点表征$\{\boldsymbol{q}_i \mid i\in V\}$，如式(1)所示。

$$\{\boldsymbol{q}_i \mid i \in V\}=\text{Structure-AwareEnhancer}(G) \tag{1}$$

与 AD-GCL[19]的数据增强模块不同，为了使得到的增强图保留更多有用信息，就需要得到准确的边表征，本文将 GATv2[20]作为结构感知增强器计算原始图中每个节点的表征。

将 $e=(u,z)$定义为图 G 中节点 u 与节点 z 的边，利用 GAT 增强器输出的节点表征得到边的权重 ω_e，如式(2)所示。

$$\omega_e=\text{MLP}([\boldsymbol{q}_u;\boldsymbol{q}_z]) \tag{2}$$

在图模型统计理论研究中，刻画边存在的概率时通常采用指数随机图模型[21]。因此，本文用指数随机图模型来刻画边 e 是否存在，将 ω_e 视作指数随机图模型的参数。每条边 $e\in E$ 将与一个随机变量 p_e 相关联，并利用 Gumbel-Max 再参数化方法[22]，使 $P(p_e=1)=e^{\omega_e}/(1+e^{\omega_e})$，$P(p_e=0)=1/(1+e^{\omega_e})$。如果 $p_e=1$，则将边保留在图 $t(G)$中，$p_e=0$ 则边 e 不存在。

3.2 对抗对比学习

3.2.1 图神经网络

对于图数据，GNN 可以对直接连接的节点进行运算，通过多层 GNN 的消息传递，图中的每个节点都可获得更多的全局信息。给定一个新闻事件的传播图 $G=(V,E)$，对于节点 i，其在第 l 层的隐藏状态表示为 $\boldsymbol{h}_i^{(l)}$，如式(3)所示。

$$\boldsymbol{h}_i^{(l)}=\text{GNN}(\boldsymbol{h}_i^{(l-1)}) \tag{3}$$

本文通过文本特征来获取初始节点表示 $\boldsymbol{h}_j^0$。令 $\boldsymbol{h}_j^0=\boldsymbol{x}_{i,j}$，$\boldsymbol{x}_{i,j}$由 TF-IDF 方法提取文本的前 d 个单词所构建的词袋模型得到，其中，d 取值为 5 000。

经过 L 层迭代后，将最终的节点表示集合进行 READOUT 操作得到图表征，如式(4)所示。

$$f(\cdot):\triangleq \boldsymbol{h}_i=\text{READOUT}(\{\boldsymbol{h}_i^L \mid i\in V\}) \tag{4}$$

这里选择了求和操作作为 READOUT 函数。此外，对传播图 G 对应的增强图 $t(G)$也进行相同的操作。

3.2.2 图对抗对比学习

在 RDEA 模型[11]中，进行图对比学习的目标是最大化虚假新闻传播图数据集的互信息(MI)，但其使用的是随机产生的图增强数据，会造成模型学习到一些多余的关系信息。例如，用户们对新闻发表看法时，会出现错位评论或乱码等现象，使传播图中产生不可靠的关系，导致模型在进行检测时效果不佳，说明这种冗余信息对虚假新闻检测是一个潜在危害。

为了解决上述问题，本文将图对抗对比学习运用到虚假新闻检测任务中。目的是优化 GNN，使原始图 G 与其增强图 $t(G)$之间的互信息最大化。同时优化图数据增强模块 $T(G)$，通过采样 $T(G)$得到 $t(G)$以最小化互信息。这样可以提高模型的对抗能力，捕获新闻传播图中少量但充分的信息，提升虚假新闻检测的效果。

此外，一个合适的图数据增强方法应该保留一定数量与新闻检测任务相关的信息。在图数据增强族$\mathcal{T}$中的图数据增强方法不应执行非常严重的扰动。因此，对于图 G 的增强图 $t(G)$，本文将 $\sum_{e\in E}\omega_e/|E|$ 添加到目标函数中，约束每个图的边扰动比例。其中，ω_e 是式(2)中边 e 扰动的概率。目标函数如式(5)所示。

$$\min_{T(\cdot)}\max_{f(\cdot)} I(f(G);f(t(G)))+\lambda_{\text{reg}}\,\mathbb{E}_G\left[\sum_{e\in E}\frac{\omega_e}{|E|}\right] \tag{5}$$

其中，$t(G)\sim T(G)$，$I(f(G);f(t(G)))$为互信息。

对于互信息 $I(f(G);f(t(G)))$，本文采用 InfoNCE[23]作为估计函数。具体来说，在训练过程中，给定一个含有 m 个图的 batch，表示为$\{G_i\}_{i=1}^m$，对于图 G_i，新闻表征 $\boldsymbol{h}_{i,1}=f(G_i)$是通过 3.2.1 节的 GNN 输出得到的表征。同样地，对于增强图$t(G_i)$，$\boldsymbol{h}_{i,2}=f(t(G_i))$，互信息项表示如式(6)所示。

$$I(f(G);f(t(G))) = \frac{1}{m}\sum_{i=1}^{m}\log\frac{\exp(\text{sim}(\boldsymbol{h}_{i,1},\boldsymbol{h}_{i,2}))}{\sum_{i'\neq i}^{m}\exp(\text{sim}(\boldsymbol{h}_{i,1},\boldsymbol{h}_{i',2}))} \tag{6}$$

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

其中，sim(·,·)表示余弦相似度。

本文将图数据增强和对抗对比学习视为进行虚假新闻检测前的预训练模块。在进行图数据增强和对抗对比学习后，通过式(4)得到新闻的预训练向量，作为下游任务的输入继续进行新闻检测。

3.3　图表示增强

3.3.1　特征增强器

当GNN层数叠加过多，节点的表示会倾向于收敛到某个值，出现过度平滑的现象[24]，即不同的新闻的表征相差很小，这样会导致虚假新闻的表征在潜在空间中更接近真实新闻表征，最后进行新闻检测时无法对新闻正确分类。为了使不同类型的新闻之间的表征差别更大，就需要考虑新闻中所有特征之间的互相影响。因此，本文利用特征增强器挖掘每个新闻事件的不变特征。该特征增强器将GNN得到的新闻表征映射到高维空间中，为模型找到每个新闻最独特的特征表示。这些特征将辅助模型进行有效的虚假新闻检测。将GNN输出的表征 $\boldsymbol{h}_i$ 通过特征增强器后得到表征 $\boldsymbol{z}_i$，如式(7)所示。

$$\boldsymbol{z}_i = g(\boldsymbol{W}_2(\max(\boldsymbol{W}_1\boldsymbol{h}_i + \boldsymbol{b}_1, \boldsymbol{0})) + \boldsymbol{b}_2) \tag{7}$$

其中，$g(\cdot)$是Dropout和标准化操作，$\boldsymbol{W}_1$ 和 $\boldsymbol{W}_2$ 为权重矩阵，$\boldsymbol{b}_1$ 和 $\boldsymbol{b}_2$ 为偏置项。

3.3.2　图表示对比学习

在对新闻传播图进行对抗对比学习后，得到输入事件图 G_i 的训练向量 $\boldsymbol{z}_i$。对于新闻事件 C_i，本文通过对新闻和所有相关评论的原始特征取平均值，得到文本图向量，即 $\boldsymbol{o}_i = \frac{1}{|V_i|}\left(\sum_{j=1}^{|V_i|-1}\boldsymbol{x}_{i,j} + \boldsymbol{x}_{i,0}\right)$。为了强调原始新闻的重要性，本文重复使用新闻文本特征 $\boldsymbol{x}_{i,0}$，然后将向量 $\boldsymbol{z}_i$、文本图向量 $\boldsymbol{o}_i$ 和新闻文本向量 $\boldsymbol{x}_{i,0}$ 进行拼接，得到最终的表征，如式(8)所示。

$$\boldsymbol{m}_i = [\boldsymbol{z}_i; \boldsymbol{o}_i; \boldsymbol{x}_{i,0}] \tag{8}$$

其中，“;”为拼接操作。

在进行新闻检测时，本文运用交叉熵损失函数对模型进行优化。由于交叉熵损失采用类间竞争机制学习类间的信息，只关心对于正确标签预测概率的准确性，忽略了其他非正确标签的差异，从而导致次优泛化和不稳定的问题[25]。而良好的泛化需要捕捉一个类中样本之间的相似性，并将类内样本与其他类中的样本进行对比。因此，本文在构造交叉熵损失函数之前引入了监督对比学习，来提高新闻表征的质量和虚假新闻检测的泛化能力。给定一个Batch B 中新闻的表征 $\{\boldsymbol{m}_i\}_{i=1}^{N_b}$（$N_b$ 为Batch的大小），当设定新闻事件 C_i 的表征 $\boldsymbol{m}_i$ 为锚点时，将具有相同标签的新闻表征 $\boldsymbol{m}_i, \boldsymbol{m}_j \in \mathcal{B}$ 作为正样本对，即 $y_i = y_j$。其中，y_i 和 y_j 分别是 $\boldsymbol{m}_i$ 和 $\boldsymbol{m}_j$ 的标签，而样本 $\{m_k \in B, k \neq i\}$ 则作为与锚点相关的负样本。然后计算对比损失，如式(9)所示。

$$L_{\text{sup}} = -\frac{1}{N_b}\sum_{\boldsymbol{m}_i \in B} l^s(\boldsymbol{m}_i) \tag{9}$$

$$l^s(\boldsymbol{m}_i) = \log \frac{\sum_{j \in B/i} \mathbb{I}_{[y_i = y_j]} \exp\left(\frac{f(\boldsymbol{m}_i, \boldsymbol{m}_j)}{\tau_s}\right)}{\sum_{j \in B/i} \exp\left(\frac{f(\boldsymbol{m}_i, \boldsymbol{m}_j)}{\tau_s}\right)} \tag{10}$$

其中，$\mathbb{I}_{[y_i = y_j]} \in \{0,1\}$ 是一个示性函数，当 $y_i = y_j$ 时取值为1，否则取值为0。τ_s 表示温度系数。

3.3.3　新闻检测

将拼接后的新闻表征 $\boldsymbol{m}_i$ 输入到全连接层和Softmax层中进行标签预测，如式(11)所示。

$$\hat{\boldsymbol{y}}_i = \text{Softmax}(\boldsymbol{W}_c\boldsymbol{m}_i + \boldsymbol{b}_c) \tag{11}$$

其中，$\boldsymbol{W}_c$ 和 $\boldsymbol{b}_c$ 为可训练参数。

对于新闻检测，本文利用交叉熵损失作为目标函数，如式(12)所示。

$$L_{\text{news}} = -\sum_{i=1}^{N_b} \boldsymbol{y}_i \log \hat{\boldsymbol{y}}_i \tag{12}$$

最终的损失函数包含两个部分：交叉熵损失和对比学习损失，计算如式(13)所示。

$$L = L_{\text{news}} + \gamma_s L_{\text{sup}} + \lambda \left\| \Theta \right\|_2^2 \tag{13}$$

其中，γ_s 为可调节超参数，$\left\| \Theta \right\|_2^2$ 表示模型的所有可训练参数，λ 表示 $L2$ 正则化系数。

4　实验分析

4.1　数据集

本文使用了两个公开数据集[6]进行评估，即Twitter15和Twitter16，数据集汇总如表1所示。

表1　数据集的统计信息

	新闻数量	#N	#F	#U	#T	平均评论数/新闻
Twitter15	1 490	374	370	374	372	223
Twitter16	818	205	205	203	205	251

Twitter15和Twitter16数据集分别包含四个不同的标签，即非虚假新闻(N)、虚假新闻(F)、未证

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

实新闻(U)和真实新闻(T)。其中,真实新闻(T)是经过专家证实是真实的辟谣新闻。

4.2 实验对比模型

为验证本文提出的模型,将 AGECL 与目前一些最先进的基线方法进行比较。

SVM-TS 模型[26]:一个基于虚假新闻生命周期的时间序列模型,利用时间序列建模技术来捕获广泛的社会背景信息。

DTC 模型[27]:一个基于监督学习的决策树模型,从每个标注的主题中提取相关特征来构建分类器,自动判断一个主题是否对应有价值的信息,并评估新闻的真实性。

RvNN 模型[7]:一种递归神经网络,深度集成了结构和内容语义信息,并利用自下而上和自上而下的树结构进行虚假新闻检测。

PPC_RNN+CNN 模型[28]:该模型结合了递归网络和卷积网络,分别捕捉用户在传播路径上的全局和局部特征变化,检测虚假新闻。

Bi-GCN 模型[9]:该模型利用自上而下的定向新闻传播图了解虚假新闻传播的模式;还利用具有相反方向的虚假新闻扩散有向图,捕捉虚假新闻扩散的结构。

RDEA 模型[11]:该模型集成了三种随机性图增强策略,通过修改图结构来提取新闻传播模式,并引入了自监督对比学习辅助虚假新闻检测任务。

4.3 参数设置

本文使用准确度(Acc)作为所有数据集的整体评估指标,利用 F_1 分数(F_1)作为每个类型的评估指标。采用随机梯度下降法更新模型参数,使用 Adam 算法对模型进行优化。使用 TF-IDF 提取文本的前 d 个单词来构建词袋模型作为文本初始特征,其中,d 的大小为 5 000。隐式特征向量的维数设置为 64,式(5)中的参数 λ_{reg} 设置为 3,式(9)中的 Dropout 概率为 0.5,式(13)中的 γ_s 为 0.6,λ 为 1e-4。本文将数据随机分成 5 个部分,并进行 5 次交叉验证,以获得稳健的结果。

4.4 结果比较

本文在两个数据集上与基线方法进行比较。表 2 分别展示了 AGECL 和所有基线在 Twitter15 和 Twitter16 数据集上的性能。

表 2 在 Twitter15 和 Twitter16 数据集上的检测结果

(单位:%)

方法	ACC	F_1			
		N	F	T	U
数据集:Twitter15					
SVM-TS	54.4	79.6	47.2	40.4	48.3
DTC	45.4	73.3	35.5	31.7	41.5
RvNN	72.3	68.2	75.9	82.1	65.4
PPC_RNN+CNN	69.7	68.9	76.0	69.6	64.5
Bi-GCN	83.6	79.1	84.2	88.7	80.1
RDEA	85.5	83.1	85.7	90.3	81.6
AGECL	**86.1**	**84.8**	**86.0**	**90.6**	**82.2**
数据集:Twitter16					
SVM-TS	57.4	75.5	42.0	57.1	52.6
DTC	46.5	64.3	39.3	41.9	40.3
RvNN	73.7	66.2	74.3	83.5	70.8
PPC_RNN+CNN	70.2	60.8	71.1	81.6	66.4
Bi-GCN	86.4	78.8	85.9	93.2	86.4
RDEA	88.0	82.3	87.8	93.7	87.5
AGECL	**88.9**	**83.5**	**89.6**	**95.0**	**87.8**

首先,所有基于深度学习的方法(RvNN, PPC_RNN+CNN, Bi-GCN 和 RDEA)性能明显优于使用手工特征的方法(SVM-TS,DTC),因为深度学习能够学习到新闻的更高级表示,以此捕获有效的特征。这展现出深度学习方法在虚假新闻检测中的重要性和优越性。

其次,AGECL 性能优于其他深度学习方法,表明了 AGECL 的有效性。在深度学习方法中,RvNN 只使用了所有叶子节点的隐藏特征向量,受最新评论的影响比较大,丢失了之前评论的信息。PPC_RNN+CNN 只把传播结构看成是平稳的时间序列,导致结构信息丢失。Bi-GCN 将新闻与评论之间的边、评论与评论之间的边都视为可靠边,这样会引入不相关的特征,训练模型时容易受到噪声交互的影响。与目前最先进的模型 RDEA 相比,AGECL 的性能更好,在 Twitter15 和 Twitter16 数据集上,AGECL 在 ACC 指标分别实现了 0.7%和 1%的整体性能提升。在四个类别(即 N、F、T、U)上的 F_1 也得到了稳步提升,例如,Twitter15 和 Twitter16 在非虚假新闻(N)类别下,F_1 分别提高

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

了 2%和 1.5%。RDEA 整合三种随机性图增强策略提取新的传播图，将这三种图增强数据作为原始新闻的正样本进行对比学习预训练。但在图增强策略中，忽略了随机删边、随机删除节点和随机选取子图会出现保留不可靠关系的现象，从而导致模型检测的准确度受到影响。不同于 RDEA，本文模型考虑了传播图结构中的不确定性，利用对抗对比学习优化图数据增强策略，使模型捕获有用信息。实验结果也证实了将对抗对比学习进行预训练的优势。

4.5　消融实验

本节进行消融实验，分析每个组件在 AGECL 中的作用。具体有以下变体。

w/o 特征增强器　移除特征增强器，使模型失去了增强的特征向量的能力。

w/o 图表示对比学习　将图表示对比学习移除，使模型在进行新闻检测时只进行交叉熵损失函数的计算。

w/o 图对抗对比学习　移除图对抗对比学习模块，对原始图和增强图进行 GNN 操作后进行无监督对比学习，并将得到的表征与原始文本信息进行融合作为新闻表征。

实验结果如表 3 所示，可以得到以下结论：①与 AGECL 模型相比，所有移除了某个模块的变体检测性能都有所下降，表明每个模块都对虚假新闻检测起到了积极作用。②特征增强器的缺失会降低 AGECL 的整体性能。因为特征增强器能将原始节点表示转化为不同的表示，有助于接下来进行的图表示对比学习使新闻检测关注到不同方面的信息。③移除图表示对比学习也降低了所有评估指标和所有数据集的性能，表明基于标签信息进行对比学习可以提高新闻表征的质量。④移除了图对抗对比学习模块后，在两个数据集上的检测准确度均有明显的下降，说明该模块的存在能降低噪声信息对新闻检测性能的影响。

表 3　对 AGECL 进行的消融实验

（单位：%）

方法	Twitter15		Twitter16	
	Acc	—	Acc	—
AGECL	**86.1**	—	**88.9**	—
w/o 特征增强器	85.8	(↓0.3)	88.5	(↓0.4)
w/o 图表示对比学习	85.9	(↓0.2)	88.7	(↓0.2)
w/o 图对抗对比学习	85.1	(↓1.0)	87.6	(↓1.3)

4.6　参数敏感性分析

我们进一步评估了模型中关键参数的影响。

• **λ_s 的影响**　λ_s 是式(13)中用于平衡目标函数中图表示对比学习任务的贡献。对于参数 λ_s，本文以 0.1 的间隔将它从 0.1 变化到 0.8。图 3(a)展示了使用原型视图对比学习对新闻表示进行建模的影响。可以观察到，当 λ_s 逐渐增大时，AGECL 的性能有所提高，并且在 $\lambda_s=0.6$ 时达到最佳性能。如果进一步增加 λ_s，模型性能开始下降。分析结果表明，图表示对比学习优化新闻表征对于本文模型的性能提升有很大贡献，图表示对比学习约束 L_{sup} 能在一定程度上进一步提高本文提出模型的表达能力。

• **λ_{reg} 的影响**　λ_{reg} 是式(5)正则化项中控制边一致性训练框架效果的参数。在进行图数据增强时，λ_{reg} 控制增强图的边扰动比例，保证原始图与进

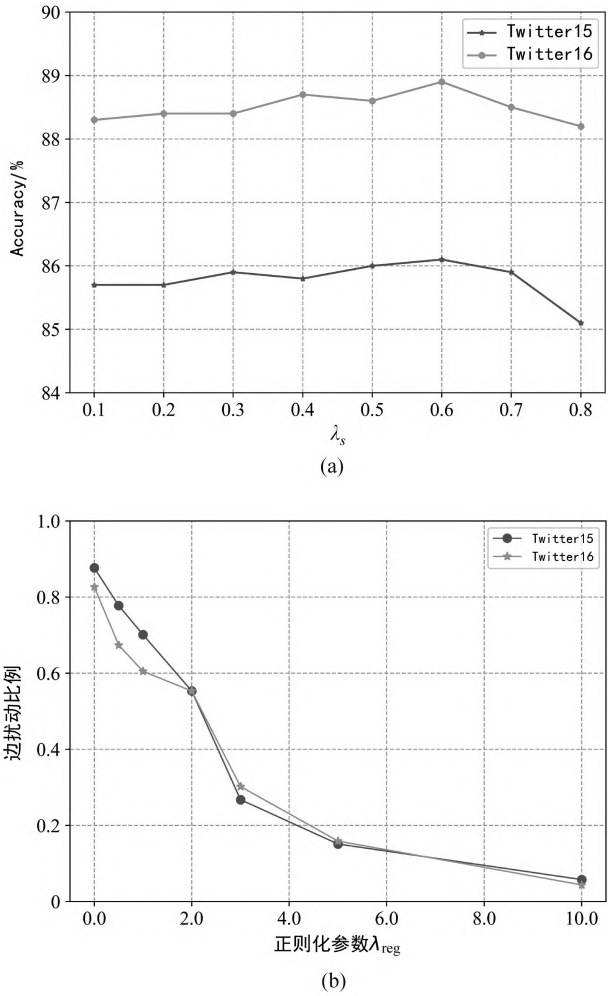


图 3　参数分析

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

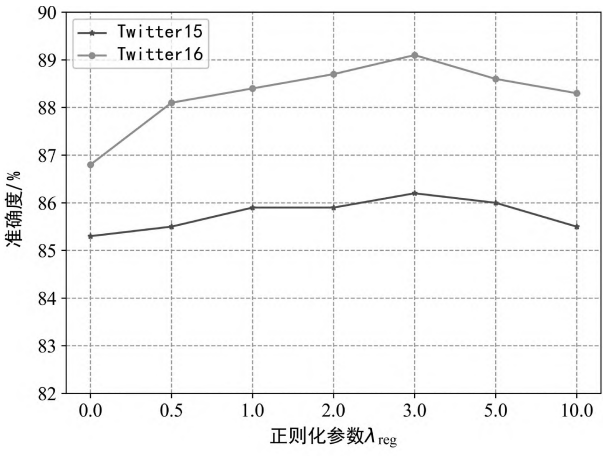


图3 （续）

行边扰动的增强图经过图对抗对比学习以后，GNN能从原始传播结构图中捕获图中最有用的信息。如图3(b)所示，λ_{reg}越大，所对应的边扰动比例$E_G\left[\sum_{e\in E}\frac{\omega_e}{|E|}\right]$越小。$\lambda_{reg}$的范围从0.0到10.0，相当于扰动80%以上的边到扰动10%以内的边。图3(c)显示了λ_{reg}在不同取值下进行虚假新闻检测的准确度得分。可以看出，当λ_{reg}取值为3.0，相当于边扰动比例在30%，Twitter15和Twitter16数据集的检测准确度达到最高。当λ_{reg}的取值范围在0.5～2.0时，即边扰动在50%～80%时，AGECL在所有数据集上仍能保持较好的性能。这说明，在一定范围内，增强图中保留的大多数是有助于分类的高质量的边。但随着边的扰动比例增大，图中边的数量减少，增强图越来越稀疏，可以学习到的信息越少，从而导致分类性能下降。

4.7 早期虚假新闻检测

为了及时预防虚假新闻的传播，在新闻传播的早期阶段进行辟谣非常重要。本文通过改变每个新闻对应的评论数量来衡量模型检测的性能，即验证该模型是否能够基于早期有限的信息量正确识别虚假新闻。

具体来说，本文使用每条新闻的前0、5、10、20、50、100条评论用于虚假新闻的早期检测。图4显示了AGECL和其他不同基线模型的性能。可以观察到，在设置最小评论数的情况下，AGECL的结果均超过了这些基线的最优结果，这表明AGECL具有优越的早期检测性能。当每条新闻只有5条评论时，AGECL在Twitter15数据集上的准确率达到85.3%；当只有10条评论时，AGECL在Twitter16数据集上的准确率达到88.6%。当输入数据只有新闻文本时，相当于新闻刚发布出来，模型的性能通常很差。这是由于缺乏评论而造成信息不足，而评论信息已被证明是对虚假新闻进行检测的关键线索。所有方法的早期性能或多或少都有波动，可能的原因是随着新闻的传播，有更多的语义和结构信息出现。同时包含噪声的评论信息也随之增加。

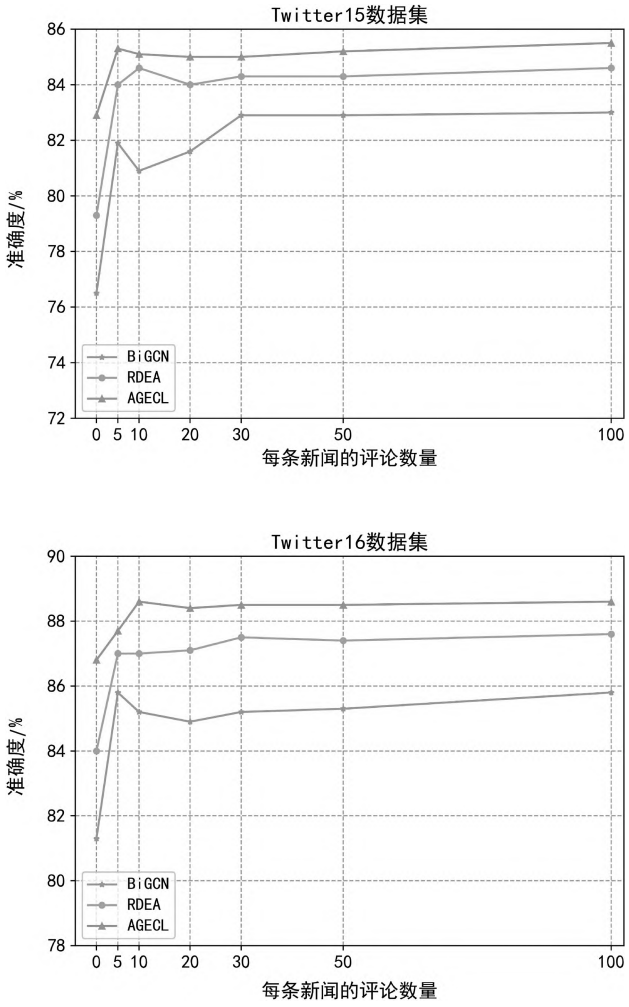


图4 早期虚假新闻检测的性能

相比之下，AGECL在所有不同评论数量下都具有稳定的性能，并超越了其他基线模型，这证明了图数据增强、对抗对比学习、图表示对比学习与传播图模型相结合可以实现稳健有效的检测。

5 总结

本文提出了一个对抗图增强对比学习模型进行虚假新闻检测。首先，采用图数据增强得到新闻增强图，使用GNN对新闻原始传播图和新闻增强图的结构信息进行编码，将输出的表征引入对抗对比学习，使该模型能够从复杂的新闻传播结构中挖掘

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

有用信息。然后通过特征增强器和图表示对比学习丰富新闻表征，提升新闻检测的效果。实验结果表明，本文模型在两个公开的真实数据集上能够有效地进行虚假新闻检测，并且在早期的虚假新闻检测任务中性能优于其他基线模型。

参考文献

[1] 陈慧敏，金思辰，林微，等. 新冠疫情相关社交媒体谣言传播量化分析[J].计算机研究与发展，2021，58(7)：1366-1384.

[2] 祖坤琳，赵铭伟，郭凯，等. 新浪微博谣言检测研究[J]. 中文信息学报，2017，31(3)：198-204.

[3] CASTILLO C，MENDOZA M，POBLETE B. Information credibility on twitter[C]//Proceedings of the 20th International Conference on World Wide Web，2011：675-684.

[4] KWON S，CHA M，JUNG K，et al. Prominent features of rumor propagation in online social media[C]//Proceedings of IEEE 13th International Conference on Data Mining. 2013：1103-1108.

[5] YANG F，LIU Y，YU X，et al. Automatic detection of rumor on sina weibo[C]//Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics，2012：1-7.

[6] 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，2016：3818-3824.

[7] MA J，GAO W，WONG K F. Rumor detection on twitter with tree-structured recursive neural networks[C]//Proceedings of Association for Computational Linguistics，2018.

[8] YU F，LIU Q，WU S，et al. A Convolutional approach for misinformation identification[C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence，2017：3901-3907.

[9] 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，2020，34(01)：549-556.

[10] 周东浩，韩文报. DiffRank：一种新型社会网络信息传播检测算法[J].计算机学报，2014，37(4)：884-893.

[11] HE Z，LI C，ZHOU F，et al. Rumor detection on social media with event augmentations[C]//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval，2021：2020-2024.

[12] VAIBHAV R M，ANNASAMY E H. HOVY. 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.

[13] YU F，LIU Q，WU S，et al. A convolutional approach for misinformation identification[C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence，2017：3901-3907.

[14] KIPF T N，WELLING M. Semi-supervised classification with graph convolutional networks[C]//Proceedings of the 5th International Conference on Learning Representations，2017.

[15] CHEN T，KORNBLITH S，NOROUZI M，et al. A simple framework for contrastive learning of visual representations[C]///Proceedings of International Conference on Machine Learning，2020：1597-1607.

[16] WU H，MA T，WU L，et al. Unsupervised reference-free summary quality evaluation via contrastive learning[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing，2020：3612-3621.

[17] CAI H，CHEN H，SONG Y，et al. Group-wise contrastive learning for neural dialogue generation[C]//Proceedings of the Association for Computational Linguistics，2020：793-802.

[18] BARBERÁ P，JOST J T，NAGLER J，et al. Tweeting from left to right：Is online political communication more than an echo chamber? [J]. Psychological Science，2015，26(10)：1531-1542.

[19] SURESH S，LI P，HAO C，et al. Adversarial graph augmentation to improve graph contrastive learning[J]. Advances in Neural Information Processing Systems，2021，34：15920-15933.

[20] BRODY S，ALON U，YAHAV E. How attentive are graph attention networks? [C]//Proceeding of the 10th International Conference on Learning Representations，2022.

[21] ROBINS G，PATTISON P，KALISH Y，et al. An introduction to exponential random graph (p*) models for social Networks[J]. Social Networks，2007，29(2)：173-191.

[22] JANG E，GU S，POOLE B. Categorical reparameterization with gumbel-softmax[C]//Proceedings of the 5th International Conference on Learning Representations，2017.

[23] POOLE B，OZAIR S，VAN DEN OORD A，et al. On variational bounds of mutual information[C]//Proceedings of International Conference on Machine Learning，2019：5171-5180.

[24] ZHAO L，AKOGLU L. Pairnorm：Tackling oversmoothing in GNNS[C]//Proceeding of the 8th International Conference on Learning Representations，2020.

[25] LIU W，WEN Y，YU Z，et al. Large-margin softmax loss for convolutional neural networks[C]//Proceedings of the 33rd International Conference on Machine Learning，2016.

[26] 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，2015：1751-1754.

[27] CASTILLO C，MENDOZA M，POBLETE B. Information credibility on twitter[C]//Proceedings of the

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

20th International Conference on World Wide Web, 2011: 675-684.
[28] 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, 2018.

陈卓敏(1998—),硕士研究生,主要研究领域为数据挖掘和自然语言处理。
E-mail: chenzhuomin1266@link.tyut.edu.cn

王莉(1971—),通信作者,博士,教授,主要研究领域为大数据计算与分析、知识图谱、数据挖掘、人工智能等。
E-mail: wangli@tyut.edu.cn

朱小飞(1979—),博士,教授,主要研究领域为自然语言处理、数据挖掘与信息检索。
E-mail: zxf@cqut.edu.cn

第二十九届全国信息检索学术会议(CCIR2023)征稿通知

信息检索旨在满足人类在互联网上快速准确地获取信息与知识的需求,研究成果将支撑国家战略决策,推动互联网和人工智能领域的发展,提升整个社会的生产效率,并对社会生活各个领域产生重大影响。全国信息检索学术会议(CCIR)由中国中文信息学会(CIPS) 举办,一路伴随着中国互联网产业的成长,是信息检索领域的旗舰会议。

第二十九届全国信息检索学术会议(The 29th China Conference on Information Retrieval, CCIR 2023)将于2023 年 11 月 23-25 在北京举行,此次会议由中国中文信息学会主办,由中国中文信息学会信息检索专委会、清华大学承办。本次会议与首届 ACM SIGIR-AP (Information Retrieval in the Asia Pacific) 会议联合举办。

征稿要求

本次会议只接收中文论文,将接收原创的有关信息检索方向的理念、算法、系统、标准与评估等方面的学术论文。本次会议不接收主体内容由人工智能模型自动生成的论文。

征文范围

论文包括但不限于以下内容:

- 搜索与排序,包括查询分析、WEB 检索、检索模型、有效性和可扩展性、信息检索理论等;
- 推荐系统,包括过滤算法、内容分析、知识获取等;
- 检索和推荐中的机器学习与自然语言处理,包括智能问答、对话系统、语义理解、知识表征和推理等;
- 人机交互,包括用户建模、交互式检索、社交媒体检索、隐私安全、用户行为分析等;
- 信息检索中的度量和评估,包括以用户为中心的评估、以系统为中心的评估等;
- 信息检索中的公平性、可靠性、透明性、可解释性;
- 垂直领域中的应用,包括移动搜索、社交搜索、结构化数据搜索、多模态搜索;其他领域的检索包括医疗、司法、教育等;
- 信息检索与前沿技术的交叉探索,包括大语言模型、脑科学等;
- 其他和信息检索相关的研究。

征文要求具体信息可登录中国中文信息学会官方公众号查阅

(中国中文信息学会)

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net