基于噪声过滤与特征增强的图神经网络欺诈检测方法

李康和[1],黄震华[1,2]
(1. 华南师范大学人工智能学院,广东佛山 528225;2. 华南师范大学计算机学院,广东广州 510631)

摘 要: 现有的基于图神经网络(Graph Neural Network,GNN)的欺诈检测方法还存在三个方面的不足:(1)没有充分考虑到样本标签分布不平衡的问题;(2)没有考虑欺诈者为了躲避检测器的检测,故意制造噪声干扰检测的问题;(3)没有考虑欺诈类型数据联系稀疏问题. 为此,本文提出一种基于噪声过滤与特征增强的图神经网络欺诈检测方法NFE-GNN(Noise Filtering and feature Enhancement based Graph Neural Network method for fraud detection)来改善欺诈检测性能. 该方法首先基于数据集的欺诈率对样本进行平衡采样;在此基础上,采用一个参数化距离函数计算节点间的相似度,并通过强化学习得到最优的噪声过滤阈值;最后,通过创建欺诈样本间的联系,丰富拓扑信息,以达到增强欺诈类特征嵌入表示的目的. 在两个公开数据集上的实验结果表明,本文所提NFE-GNN方法的性能优于目前主流的图神经网络欺诈检测方法.

关键词: 欺诈检测;类不平衡;节点分类;图结构数据;图神经网络;性能评估
基金项目: 国家自然科学基金(No.62172166);广东省基础与应用基础研究基金(No.2022A1515011380)
中图分类号: TP311.5;TP391.4 **文献标识码:** A **文章编号:** 0372-2112(2023)11-3053-08
电子学报 URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20230489

Noise Filtering and Feature Enhancement Based Graph Neural Network Method for Fraud Detection

LI Kang-he[1], HUANG Zhen-hua[1,2]
(1. *School of Artificial Intelligence*, *South China Normal University*, *Foshan*, *Guangdong* 528225, *China*;
2. *School of Computer Science*, *South China Normal University*, *Guangzhou*, *Guangdong* 510631, *China*)

Abstract: Existing graph neural network (GNN)-based fraud detection methods have at least three shortcomings: (1) They do not adequately consider the problem of imbalanced distribution of sample labels. (2) They do not take into account the problem that fraudsters deliberately create noise to interfere with fraud detection in order to avoid detection by detectors. (3) They fail to consider the limitations of sparse connections for fraud data. To address these three shortcomings, this paper proposes a fraud detection method, called NFE-GNN (Noise Filtering and feature Enhancement based Graph Neural Network method for fraud detection), to improve the fraud detection performance. The proposed NFE-GNN method first employs a dataset-based fraud rate sampling technology to achieve a balance of benign and fraudulent samples. Based on this, a parameterized distance function is introduced to calculate the similarities between nodes, and the optimal noise filtering threshold is obtained through adaptive reinforcement learning. Finally, an effective algorithm is presented to increase the connections between fraudulent samples, and enrich the topology information in the graph to enhance the feature representation capability of fraudulent samples. The experimental results on two publicly available datasets demonstrate that the detection performance of the proposed NFE-GNN method is better than that of state-of-the-art graph neural network methods.

Key words: fraud detection; class imbalance; node classification; graph data; graph neural network; performance evaluation
Foundation Item(s): National Natural Science Foundation of China (No.62172166); Guangdong Basic and Applied Basic Research Foundation (No.2022A1515011380)

收稿日期:2023-06-01;修回日期:2023-09-23;责任编辑:覃怀银

1 引言

欺诈检测通过利用用户的行为模式或者交易模式数据进行模型训练,进而生成检测模型,来预测某一用户或者交易是否存在欺诈行为. 目前,在诸如金融[1]、医疗[2]、通信[3]、新闻[4]和评论管理[5,6]等领域都有广泛的应用.

图神经网络(Graph Neural Network,GNN)具有挖掘实体之间深刻的依赖性和相关性的能力,所以基于图神经网络的欺诈检测成为新的主流方法. 在传统的图神经网络[7~9]基础上,Shi等[10]提出DR-GCN方法. Zeng等[11]提出GraphSAINT方法,能够有效应用于大部分的欺诈场景. 然而,在噪声信息过多的情况下,检测性能将会显著下降. Li等[12]提出GAS模型,缓解了特征不一致和关系不一致的问题. Liu等[5]在同构图上进一步改进了检测方法. Dou等[6]针对欺诈者的伪装特性,提出了CARE-GNN方法,在此基础上,Liu等[13]提出了HA-GNN方法. 针对图的不平衡问题以及不一致问题,Zhang等[14]提出了FRAUDRE方法. Liu等[15]提出PC-GNN方法进一步改善节点的采样. 此外,Huang等[16]提出AO-GNN方法,采用最大化AUC(Area Under Curve,AUC)来解决标签不平衡问题,并在欺诈检测任务上达到了最佳性能.

我们分析发现,上述方法[11~16]虽然能充分利用图数据的拓扑结构信息,提高欺诈检测的性能,但仍然存在三个方面的不足:

(1)对数据进行采样时,没有充分利用数据中的标签信息,从而导致采样后,数据中良性类别和欺诈类别样本不仅没到达到平衡状态,甚至可能造成关键拓扑信息丢失. 这将导致在进行目标节点领域信息聚合的过程中,特征嵌入表示不理想.

(2)在进行欺诈检测时,没有充分考虑欺诈者为躲避检测器的检测,故意制造噪声信息的情况,这将使得检测方法对节点进行特征嵌入表示学习时,欺诈节点特征会掺杂很多噪声信息,导致学习到的良性类节点特征和欺诈类节点特征之间区分度不高,进一步造成检测效果不理想的结果.

(3)没能有效解决欺诈类数据联系稀疏的问题,导致欺诈类特征难以得到学习. 例如,在电子商务购物应用场景中,欺诈者会为了避免彼此一起被检测出来,通常会避免直接发生相互联系.

为了解决以上不足,本文提出了一种基于噪声过滤与特征增强的图神经网络欺诈检测方法NFE-GNN(Noise Filtering and feature Enhancement based Graph Neural Network method for fraud detection).

2 符号与定义

本节主要给出本文中使用的符号与定义.

定义1 多关系无向图:给定一个无向图$G=\{V,R,E,\boldsymbol{A},X,Y\}$,其中,$V=\{v_1,v_2,\cdots,v_n\}$为$G$中的$n$个节点构成的集合;关系集合$R$包含$\delta$个关系,即$R=\{r_1,r_2,\cdots,r_\delta\}$;$E=\{\varepsilon_{r1},\varepsilon_{r2},\cdots,\varepsilon_{r\delta}\}$表示$G$中所有边构成的集合,而$\varepsilon_{ri}(1\leqslant i\leqslant\delta)$为关系$r_i$对应边所构成的集合,$\boldsymbol{A}$表示为节点的领接矩阵,在$R$关系集合中,节点间不同关系对应的领接矩阵表示为$\boldsymbol{A}_{ri}\in\{\boldsymbol{A}_{r1},\boldsymbol{A}_{r2},\cdots,\boldsymbol{A}_{r\delta}\}(1\leqslant i\leqslant\delta)$;$X=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\}$表示$V$中节点的特征集合,每个节点$v_j(1\leqslant j\leqslant n)$有一个$m$维的特征向量$\boldsymbol{x}_j\in X$,$Y$表示为$V$中节点的标签集合.

定义2 图神经网络欺诈检测[6]:基于定义1中的多关系无向图G,目标节点v在GNN第l层以及关系r下的特征表示可定义为

$$\boldsymbol{h}_{v,r}^{(l)}=\sigma(\boldsymbol{W}_r^{(l)}(\boldsymbol{h}_{v,r}^{(l-1)}\oplus A_r^{(l)}\{\boldsymbol{h}_{v',r}^{(l-1)}\})) \tag{1}$$

其中,$\boldsymbol{h}_{v,r}^{(l-1)}$为节点$v$在GNN第$l-1$层以及关系$r$下的特征表示,$\boldsymbol{h}_{v',r}^{(l-1)}$表示$v$的邻居节点$v'$在第$l-1$层以及关系$r$下的特征表示,$\boldsymbol{W}_r^{(l)}$为权重参数,$A_r^{(l)}\{\}$为不同关系下的领域信息聚合函数.$\sigma$为激活函数,"⊕"表示组合节点$v$的信息及其领域信息的操作,例如,拼接或求和. 在此基础上,将节点v的第$l-1$层特征表示$\boldsymbol{h}_v^{(l-1)}$,及其在第l层中各关系下的特征表示进行融合,从而得到最终的特征表示:

$$\boldsymbol{h}_v^{(l)}=\sigma(\boldsymbol{W}^{(l)}(\boldsymbol{h}_v^{(l-1)}\oplus\boldsymbol{h}_{v,r1}^{(l)}\oplus\cdots\oplus\boldsymbol{h}_{v,r\delta}^{(l)})) \tag{2}$$

其中,$r_1,r_2,\cdots,r_\delta\in R$,$\boldsymbol{W}^{(l)}$为权重参数.

3 NFE-GNN方法

本文提出的NFE-GNN方法的总体架构如图1所示,该方法主要由四个模块组成:采样模块、噪声过滤模块、特征增强与信息聚合模块.

3.1 标签平衡采样

根据想法需要,我们构造一个多关系无向图$G=\{V,R,E,\boldsymbol{A},X,Y\}$,对于节点$v\in V$,根据它所属的类型,其被采样的权重定义如下:

$$f_v=\begin{cases}|V|, & v\in 欺诈类\\ a\cdot\dfrac{|V|}{\gamma}, & v\in 良性类\end{cases} \tag{3}$$

式(3)中,$|V|$为V中的节点数量,γ表示样本中欺诈率的大小,a为超参数.f_v越大,节点v被采样的概率就越小,这样就可以根据数据集的欺诈率的大小,对欺诈节点和良性节点进行不同概率的采样. 对于节点v,结合一阶邻居数量信息,其最终被采样的概率定义如下:

$$p_v=\frac{\sum_{i=1}^{|V|}\boldsymbol{A}_{ij}}{f_v}=\frac{\boldsymbol{H}_{jj}}{f_v} \tag{4}$$

其中,j为节点v在领接矩阵$\boldsymbol{A}$中的位置编号,$\boldsymbol{H}$为$\boldsymbol{A}$对应的度矩阵. 不难得出,$\sum_{i=1}^{|V|}\boldsymbol{A}_{ij}$和$\boldsymbol{H}_{jj}$为节点$v$一阶邻居

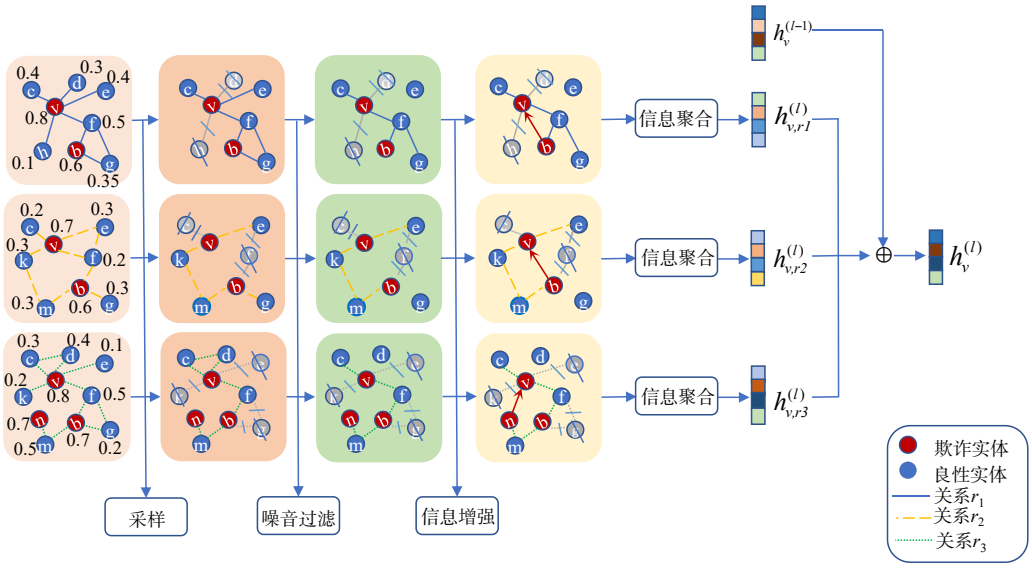


图1　NFE-GNN模型总体架构

数量. 根据式(4)可知,节点v被采样的概率与它的一阶邻居数量成正比,与其被采样权重f_v成反比. 这样就可以采集到更多属于少数类的节点. 经过采样之后的训练子图的样本标签处于相对平衡状态.

3.2　长距离噪声过滤模块

3.2.1　标签相似度测量

受CARE-GNN[6]的思想启发,本文采用目标节点v与其邻居节点v'之间基于参数化距离函数的余弦特征相似度计算方法. 具体来说,对于目标节点v,在GNN的第l层以及关系r下,其与邻居节点v'之间的余弦相似性为

$$S_r^{(l)}(v,v')=\frac{\sigma(\mathrm{MLP}(\boldsymbol{h}_v^{(l-1)}))\cdot\sigma(\mathrm{MLP}(\boldsymbol{h}_{v'}^{(l-1)}))}{\left\|\sigma(\mathrm{MLP}(\boldsymbol{h}_v^{(l-1)}))\right\|\left\|\sigma(\mathrm{MLP}(\boldsymbol{h}_{v'}^{(l-1)}))\right\|} \tag{5}$$

在此基础上,可以定义目标节点与其邻居的余弦距离为

$$D^{(l)}(v,v')=1-S_r^{(l)}(v,v') \tag{6}$$

其中,σ表示激活函数,$S^{(l)}(v,v')$表示为节点v和节点v'的特征基于参数化距离函数的余弦相似度,$D^{(l)}(v,v')$表示v和v'间的余弦距离. $D^{(l)}(v,v')$的值越小,表明v和v'之间的相似性越大,反之相反.

通过式(5)和式(6)对目标节点及其所有邻居进行余弦距离计算,并按照升序方式进行排列,排列后的邻居距离用集合$D(v')$表示,然后对这些邻居使用强化学习(Reinforcement Learning,RL)[17]得到的最优阈值,进行噪声过滤. 噪声过滤的过程可以定义如下:

$$N_r^{(l)}(v')=\{v'\in V\mid \mathrm{Top}(D(v'))\} \tag{7}$$

其中,$N_r^{(l)}(v')$表示阈值过滤后剩下的邻居集合,参与中心节点v的特征表示过程,本文此处的$\mathrm{Top}(D(v'))$表示选择集合$D(v')$的前S_v个数据,$S_v\in[0,1]$为通过强化学习找到的最优过滤阈值. 当阈值$S_v=1$时,目标节点的所有邻居都将被保留下来,参与目标节点的领域信息聚合过程;当阈值$S_v=0$时,该目标节点的所有邻居都会被过滤掉,不参与目标节点领域信息聚合的过程.

由于距离噪声过滤模块是一个神经网络,所以需要结合一个损失函数进行训练,本文使用交叉熵损失函数进行训练:

$$L_{\mathrm{sim}}^{(l)}=-\sum_{v\in V}\log(y_v\cdot\sigma(\mathrm{MLP}(\boldsymbol{h}_v^{(l)}))) \tag{8}$$

其中,y_v表示为目标节点v的标签信息,σ表示激活函数,我们此处使用tanh函数作为激活函数.

3.2.2　强化学习自适应寻找最优阈值

我们将强化学习过程分为阈值调整机制、奖励机制和状态机制三个部分:

(1)阈值调整机制:根据奖励机制选择调整噪声过滤阈值,直至找到最优阈值或者触发状态机制的终止条件. 由于阈值$S_v\in[0,1]$,本文定义一个超参数τ,阈值调整机制会根据奖励机制规则去增加或者减去一个单位τ或者减零的方式对阈值进行更新.

(2)奖励机制:根据连续两个训练回合(epoch)之间节点的平均距离的变化做出相对应改变. 节点之间每一种连接关系的节点平均距离变化可能不一样. 对于每个训练回合,在GNN的第l层以及关系r下节点的平均距离定义为

$$\overline{d}\,(D_r^{(l)})^{(e)}=\frac{\sum_{v'\in V}D_r^{(l)}(v,v')^{(e)}}{K} \tag{9}$$

其中,K表示在每个训练回合中训练样本的数量. 在此基础上,给每一个训练回合定义一个奖励如下:

$$\mathrm{RW}=\begin{cases}+, & \overline{d}(D_r^{(l)})^{(e+1)}-\overline{d}(D_r^{(l)})^{(e)}<0\\ 0, & \overline{d}(D_r^{(l)})^{(e+1)}-\overline{d}(D_r^{(l)})^{(e)}=0\\ -, & \overline{d}(D_r^{(l)})^{(e+1)}-\overline{d}(D_r^{(l)})^{(e)}>0\end{cases} \tag{10}$$

当后一个训练回合节点之间的平均距离小于前一个训练回合节点之间的平均距离时，意味着可以发现更多更相似的节点，更大的阈值可以使得选择更多相似的邻居参加中心节点的嵌入表示，同时奖励机制的奖励值则为正；当平均距离不变时，奖励值为0；否则，奖励值为负.

$$S_v=\begin{cases}S_v+\tau, & \mathrm{RW}=+\\ S_v-0, & \mathrm{RW}=0\\ S_v-\tau, & \mathrm{RW}=-\end{cases} \tag{11}$$

式(11)给出阈值更新情况，当回报值大于0时，过滤阈值S_v就会增加τ，当回报值小于0时，S_v就会减τ，否则S_v保持不变.

(3)状态机制：最优过滤阈值不再变化或达到模型设置的最大训练回合数时，执行结束.

3.3 特征增强与信息聚合

基于3.2小节中计算目标节点v与邻居节点v'之间的余弦距离方法，把它从目标节点与其邻居之间的测量，拓展到目标节点与整个图上. 当目标节点为欺诈类型时，计算其与训练样本中其余所有属于欺诈类型样本之间的余弦距离，同时对该距离进行升序排列，采用Top-k的选择方法来生成欺诈节点之间的边，创建欺诈实体之间的联系，被选择的欺诈节点参与目标节点的最终特征表示. 欺诈实体之间新创建关系的标准定义如下：

$$N_{\mathrm{new},r}^{(l)}(v)=\{v'\in V\mid \mathrm{Top}(D(v'))\},C(v')=C(v) \tag{12}$$

其中，$C(v')=C(v)$表示节点v'与目标节点v同属于相同类型，均为欺诈类. $N_{\mathrm{new},r}^{(l)}(v)$表示在GNN第$l$层以及关系$r$下被选择为目标节点$v$的新邻居，参与节点$v$的领域信息聚合过程，此处的$\mathrm{Top}(D(v'))$表示取距离集合的Top-$k$数据，从式(12)可以看出，当目标节点$v$为欺诈类型样本时，与其创建新连接的节点也必须是欺诈类型的节点.

最后就是对目标节点v进行信息聚合表示. 本文的信息聚合阶段分为单关系内信息聚合和多关系间信息融合两部分. 目标节点v在单关系r内的信息聚合可以使用第3节中的式(1)来计算，而多关系间信息融合通过式(2)来实现.

3.4 模型训练

本文NFE-GNN方法的损失函数由三部分组成，即：长距离噪声过滤模块的损失$L_{\mathrm{sim}}^{(l)}$、节点分类损失L_{cls}以及防止模型过拟合的L_2正则项：

$$L_{\mathrm{NFE-GNN}}=L_{\mathrm{cls}}+\lambda_1\sum_{l=1}^{L}L_{\mathrm{sim}}^{(l)}+\lambda_2\|\boldsymbol{\Theta}\|_2 \tag{13}$$

其中，L为GNN的总层数，$\boldsymbol{\Theta}$表示NFE-GNN方法的可学习参数，$\|\cdot\|_2$表示L_2范式，λ_1和λ_2表示各部分损失的权重值. 式(14)给出节点分类损失函数L_{cls}，为交叉熵损失函数：

$$L_{\mathrm{cls}}=-\sum_{v\in V}\log(y_v\cdot\sigma(\mathrm{MLP}(\boldsymbol{z}_v))) \tag{14}$$

其中，$\boldsymbol{z}_v=\boldsymbol{h}_v^{(L)}$，为目标节点$v$的最后一层(即第$L$层)的特征表示. $L_{\mathrm{sim}}^{(l)}$在式(8)中定义.

4 实验评估

4.1 数据集

本文实验使用图神经网络欺诈检测任务常用的两个公开数据集，即YelpChi和Amazon，表1给出具体信息(ALL表示节点间的连接不区分关系看待).

在YelpChi数据集中，每一个样本都有一个32维的原始特征，样本之间存在以下三种不同的连接关系，表示节点之间共享属性或者产生直接互动：

(1)R-U-R：在同一用户发布的评论间连边.

(2)R-S-R：在同一产品下具有相同星级的评论连边.

(3)R-T-R：在一个月内对同一产品发布的两个评论连边.

Amazon数据集的每一个样本对应一个25维的原始特征，样本间存在以下三种不同的连接关系：

(1)U-P-U：查看至少一个相同产品的用户之间连边.

(2)U-S-U：在一周内至少有一个相同星级的用户之间连边.

(3)U-V-U：在所有用户中，相互评论文本相似度最高的前5%(由TF-IDF指标度量)的用户之间连边.

表1 数据集的统计信息

数据集	节点/个	欺诈率/%	关系	边数/条
YelpChi	45 954	14.5	R-U-R	49 315
			R-T-R	573 616
			R-S-R	3 402 743
			ALL	3 846 979
Amazon	11 944	9.5	U-P-U	175 608
			U-S-U	3 566 479
			U-V-U	1 036 737
			ALL	4 398 392

4.2 比较方法

为了检验本文提出的NFE-GNN方法的有效性，我们比较了9个最新的图神经网络模型及其改进方法：GCN[7]，GAT[8]，GraphSAGE[9]，DR-GCN[10]，GraphSAINT[11]，PC-GNN[15]，GraphConsis[5]，CARE-GNN[6]以及AO-GNN[16].

4.3 评价指标和超参数设置

与现有相关工作一致，本文实验采用如下三个评

估指标:AUC、F_1-macro和G-Mean.

本文使用Adam[18]优化器对NFE-GNN方法的参数进行优化,其他超参数最优设置在表2中加粗表示,其中,d表示图神经网络的节点特征的维度,b为最小批次样本大小,lr表示学习率,k为特征增强模块中,选择相似度排名前k的欺诈样本,λ_1表示相似度损失权重,λ_2表示模型参数L_2正则化损失权重,τ为强化学习的步长,μ表示模型迭代的次数,L为模型的层数. 其他比较方法使用的参数设置都是根据原论文提供的源码进行设置,所有比较方法的训练集、验证集和测试集分别设置为40%、20%和40%.

表2　本文NFE-GNN方法的超参数设置以及调参过程的参数选择

	超参	d	b	lr	k	λ_1	λ_2	τ	μ	L
数据集	YelpChi	{8,**16**,32,64}	**1 024**	**0.01**	{1,0.5,**0.25**,0.125}	{0.5,1,**2**,4}	**0.001**	{0.08,0.06,0.04,**0.02**,0.01}	**700**	{1,2,3,4,**5**,6}
	Amazon	{8,**16**,32,64}	**256**	**0.005**	{1,0.5,**0.25**,0.125}	{0.5,1,**2**,4}	**0.001**	{0.08,0.06,0.04,**0.02**,0.01}	**700**	{1,2,3,4,**5**,6}

4.4　与现有方法的性能比较

表3给出了所有比较方法的检测性能,从中可以看出NFE-GNN方法的性能优于所有对比方法,证明了其有效性. 特别地,我们得出以下几个结论:

(1)与经典的图神经网络模型GCN、GAT和GraphSAGE相比,本文NFE-GNN方法检测性能得到显著的提升,主要原因是这些经典的图神经网络模型没有专门针对欺诈场景进行特殊设计,从而得不到最好的检测效果. GCN模型与GAT模型相比,AUC指标在YelpChi数据集上提高了超过2%,主要原因是GAT模型是注意力机制,模型参数较多,在类不平衡场景下,少数类的数据变得更加稀缺,很难得到足够的数据来充分训练模型. GraphSAGE模型在这三种经典的图神经网络中,检测性能最差,主要原因是它在对目标节点进行领域采样时,没有充分考虑欺诈场景,盲目地对目标节点的邻居进行固定数量的采样. 在欺诈检测任务中,不合适的采样可能会导致欺诈类数据变得更加少,甚至完全被过滤掉,从而致使模型检测性能变得更加不理想. 这也从侧面表明了合适的采样方式的必要性.

(2)基于类不平衡处理的图神经网络欺诈检测模型DR-GCN、GraphSAINT和PC-GNN的性能优于经典的图神经网络模型,但弱于本文的NFE-GNN方法. 由于这三个模型充分考虑到欺诈场景,进行针对性的采样,因此,检测性能比经典的图神经网络模型有较大的提升. 例如,与经典图神经网络模型中性能最好的GCN相比,GraphSAINT模型把采样的过程,分配到每一个子图上,提高了采样的效果,使得在YelpChi数据集上的AUC指标比GCN提升了10.16%;然而,该模型采样的过程缺乏考虑标签信息,导致采样之后的类别标签仍然处于不平衡状态. 基于此,PC-GNN模型进一步改善了样本采样的过程,充分利用了标签信息,使得AUC指标比GraphSAINT提高了9.88%,这进一步说明了结合标签分布的采样方式是有效的.

(3)GraphConsis、CARE-GNN和AO-GNN是针对数据集中可能存在伪装行为或者噪声而提出的模型. 与经典的图神经网络模型相比,它们检测性能有着明显的提升. 例如,GraphConsis模型考虑到数据中存在噪声信息,对上下文不一致的噪声节点进行过滤,使得它在YelpChi数据集上的AUC指标比GCN提高了10%. CARE-GNN模型在GraphConsis的基础上,进一步改善了噪声过滤方式,致使在YelpChi数据集上的AUC指标进一步提高6.36%. 之后,AO-GNN模型进一步改进了噪声过滤的方式,利用边的修剪策略对噪声进行过滤,从而,在YelpChi数据集上的AUC指标进一步提高近11.86%. 这也充分表明噪声也是欺诈检测任务中的一个负面因素.

表3　本文NFE-GNN方法与现有9个模型的检测性能比较

	数据集	YelpChi			Amazon		
	评估指标	AUC	F_1-macro	G-Mean	AUC	F_1-macro	G-Mean
现有方法	GCN	0.598 3	0.562 0	0.436 5	0.836 9	0.648 6	0.571 8
	GAT	0.571 5	0.487 9	0.165 9	0.810 2	0.646 4	0.667 5
	GraphSAGE	0.543 9	0.440 5	0.258 9	0.758 9	0.641 6	0.594 9
	DR-GCN	0.592 1	0.552 3	0.403 8	0.829 5	0.648 8	0.535 7
	GraphSAINT	0.699 9	0.596 0	0.590 8	0.870 1	0.762 6	0.796 3
	PC-GNN	0.798 7	0.630 0	0.716 0	0.958 6	0.895 6	0.903 0
	GraphConsis	0.698 3	0.587 0	0.585 7	0.874 1	0.751 2	0.767 7
	CARE-GNN	0.761 9	0.633 2	0.679 1	0.906 7	0.899 0	0.896 2
	AO-GNN	0.880 5	0.704 2	0.813 4	0.964 0	0.892 1	0.909 6
本文方法	NFE-GNN	**0.901 3**	**0.755 3**	**0.825 7**	**0.976 2**	**0.919 2**	**0.931 1**

4.5 消融实验

本小节通过消融实验来验证NFE-GNN方法三个关键模块的有效性. 我们构造3个变体方法:

(1)NFE-GNN/n: 在完整方法NFE-GNN中,只使用噪声过滤组件和特征增强组件,而不使用标签平衡采样组件.

(2)NFE-GNN/f: 在NFE-GNN方法中,只使用标签平衡采样组件和特征增强组件,而不使用噪声节点过滤组件.

(3)NFE-GNN/e: 在NFE-GNN方法中,只使用标签平衡采样组件和噪声节点过滤组件,而不使用特征增强组件.

实验结果如表4所示. 首先,从表4中可以看出,由于NFE-GNN/n模型只使用噪声过滤模块和特征增强模块,而不使用标签平衡采样模块,导致在模型训练过程中,欺诈类与良性类样本之间的差距较大,模型难以很好地学习到少数类特征. 这表明标签平衡采样模块在欺诈检测任务中是必要的.

表4 NFE-GNN模型及其3个变种模型的检测性能

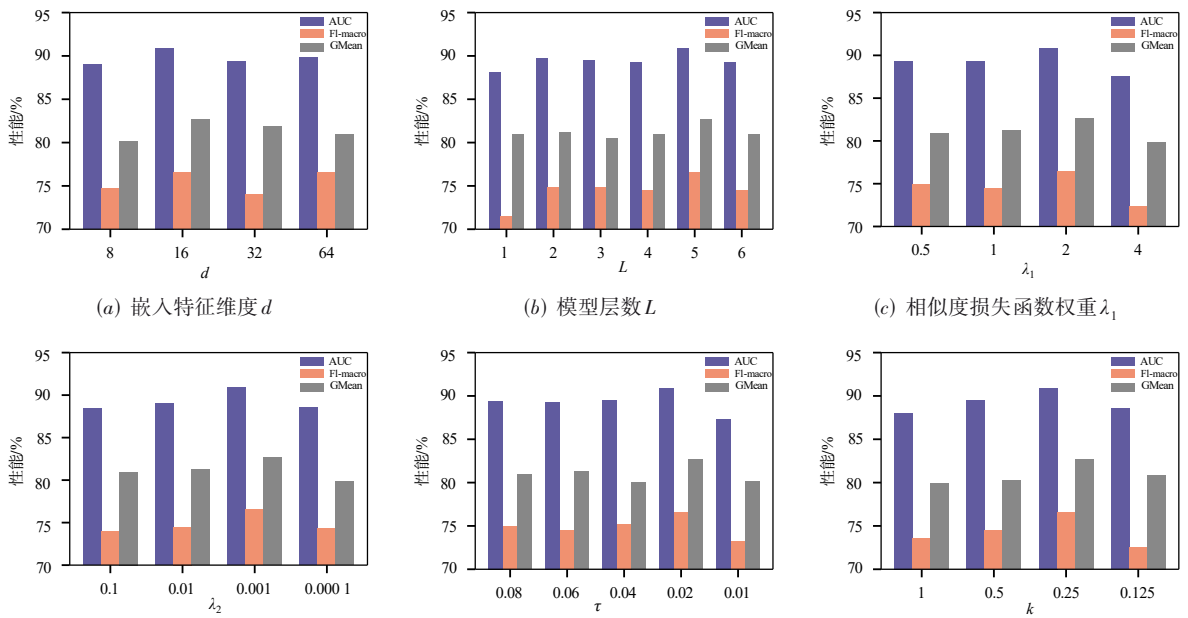
	数据集	YelpChi			Amazon		
	评估指标	AUC	F_1-macro	G-Mean	AUC	F_1-macro	G-Mean
变体方法	NFE-GNN/n	0.871 3	0.714 5	0.794 4	0.952 9	0.895 7	0.906 5
	NFE-GNN/f	0.891 2	0.747 9	0.812 1	0.968 9	0.911 2	0.920 3
	NFE-GNN/e	0.882 5	0.728 8	0.810 1	0.965 4	0.899 8	0.916 2
完整方法	NFE-GNN	**0.901 3**	**0.755 3**	**0.825 7**	**0.976 2**	**0.919 2**	**0.931 1**

其次,实验结果表明,与完整方法NFE-GNN相比,NFE-GNN/f方法在YelpChi和Amazon数据集上的三个性能评估指标均降低. 这主要是因为在现实的欺诈场景中,欺诈者通常为了逃避检测器的检测,故意制造许多噪声信息的情况,这些噪声信息也是导致检测性能不理想的原因之一.

再次,从实验结果还可以看出,由于NFE-GNN/e方法只使用标签平衡采样模块和噪声过滤模块,而不使用特征增强模块,导致模型最后学习到的欺诈类与良性类样本间的特征没有得到明显的区分,主要原因是,由于少数类之间的联系过少,能得到的少数类特征信息相当有限. 这说明了增强少数类特征能够有效的提高检测性能.

4.6 参数敏感度分析

本小节探索和分析NFE-GNN方法中6个重要超参数(表2给出的参数$d,L,\lambda_1,\lambda_2,\tau,k$)对检测性能的影响:图2($a$)~($f$)给出了YelpChi数据集上的实验结果.



(a) 嵌入特征维度d (b) 模型层数L (c) 相似度损失函数权重λ_1

(d) 参数L_2范式损失权重λ_2 (e) 强化学习步长τ (f) 欺诈类top-k相似度k值

图2 NFE-GNN方法参数敏感度分析

从图2(*a*)可以看出,增大目标节点的特征表示维度*d*,并不能持续带来更优的检测性能. 从图2(*b*)可以发现,GNN的总层数*L*从1到5时,性能评估指标AUC和G-Mean基本保持不变,F_1-macro呈现逐渐递增的趋势,当*L*=5时,三个性能指标都达到最大,而超过了5层之后,NFE-GNN方法出现了过拟合,从而导致性能下降. 图2(*c*)和图2(*d*)探索了λ_1和λ_2对检测性能的影响,从中可以看出,当$\lambda_1=2$,并且$\lambda_2=0.01$时,NFE-GNN方法的检测性能达到最佳. 图2(*e*)展示了强化学习步长τ对欺诈检测性能的影响,当$\tau=0.02$时,NFE-GNN方法的检测性能达到最大值,而过大或者过小的步长,都可能使得模型无法找到最优的噪声过滤阈值. 最后,图2(*f*)探索了在特征增强模块中,超参数*k*对欺诈检测性能的影响. 我们发现,当$k=0.25$时,NFE-GNN方法的检测性能达到最优.

综上可知,本文需要选择合适的超参数才能够使得NFE-GNN方法的检测性能达到最佳,过大或过小的超参数都会影响模型最终的检测性能.

5　结论

现有图神经网络欺诈检测方法存在类不平衡、噪声信息过多以及欺诈类之间的联系稀疏问题,造成检测性能不理想,本文提出了一种根据欺诈率采样平衡训练数据,然后通过特征相似度对噪声信息进行过滤,最后创建欺诈类之间的联系,丰富节点的拓扑信息,以增强节点的特征嵌入. 在两个公开数据集上实验验证了NFE-GNN方法的有效性.

参考文献

[1] VAN BELLE R, BAESENS B, DE WEERDT J. CATCHM: A novel network-based credit card fraud detection method using node representation learning[J]. Decision Support Systems, 2023, 164: 113866.

[2] LUO X D, WANG G T, LIAO W J, et al. Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency[J]. Medical Image Analysis, 2022, 80: 102517.

[3] NI P F, YU W. A victim-based framework for telecom fraud analysis: A Bayesian network model[J]. Computational Intelligence and Neuroscience, 2022, 2022: 1-13.

[4] CUI J, KIM K, NA S H, et al. Meta-path-based fake news detection leveraging multi-level social context information [C]//Proceedings of the 31st ACM International Conference on Information & Knowledge Management. New York: ACM, 2022: 325-334.

[5] LIU Z W, DOU Y T, YU P S, et al. Alleviating the inconsistency problem of applying graph neural network to fraud detection[C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2020: 1569-1572.

[6] DOU Y T, LIU Z W, SUN L, et al. Enhancing graph neural network-based fraud detectors against camouflaged fraudsters[C]//Proceedings of the 29th ACM International Conference on Information & Knowledge Management. New York: ACM, 2020: 315-324.

[7] 徐冰冰, 岑科廷, 黄俊杰, 等. 图卷积神经网络综述[J]. 计算机学报, 2020, 43(5): 755-780.
XU B B, CEN K T, HUANG J J, et al. A survey on graph convolutional neural network[J]. Chinese Journal of Computers, 2020, 43(5): 755-780. (in Chinese)

[8] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[C]//Proceedings of the International Conference on Learning Representations. Vancouver: ICLR, 2018: 1-12.

[9] HAMILTON W L, YING R, LESKOVEC J. Inductive representation learning on large graphs[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 1025-1035.

[10] SHI M, TANG Y F, ZHU X Q, et al. Multi-class imbalanced graph convolutional network learning[C]//Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence. Yokohama: International Joint Conferences on Artificial Intelligence Organization, 2020: 2879-2885.

[11] ZENG H, ZHOU H, SRIVASTAVA A, et al. Graphsaint: Graph sampling based inductive learning method[EB/OL]. (2020-02-16)[2023-05-24]. https://doi.org/10.48550/arXiv.1907.04931.

[12] LI A, QIN Z, LIU R S, et al. Spam review detection with graph convolutional networks[C]//Proceedings of the 28th ACM International Conference on Information and Knowledge Management. New York: ACM, 2019: 2703-2711.

[13] LIU Y J, SUN Z Y, ZHANG W S. Improving fraud detection via hierarchical attention-based graph neural network [J]. Journal of Information Security and Applications, 2023, 72: 103399.

[14] ZHANG G, WU J, YANG J, et al. FRAUDRE: Fraud detection dual-resistant to graph inconsistency and imbalance[C]//2021 IEEE International Conference on Data Mining (ICDM). Piscataway: IEEE, 2022: 867-876.

[15] LIU Y, AO X, QIN Z D, et al. Pick and choose: A GNN-

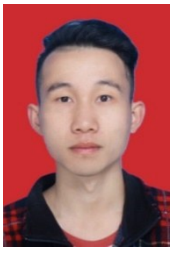
based imbalanced learning approach for fraud detection [C]//Proceedings of the Web Conference 2021. New York: ACM, 2021: 3168-3177.

[16] HUANG M D, LIU Y, AO X, et al. AUC-oriented graph neural network for fraud detection[C]//Proceedings of the ACM Web Conference 2022. New York: ACM, 2022: 1311-1321.

[17] BRUNKE L, GREEFF M, HALL A W, et al. Safe learning in robotics: From learning-based control to safe reinforcement learning[J]. Annual Review of Control, Robotics, and Autonomous Systems, 2022, 5: 411-444.

[18] KINGMA D P, BA J. Adam: A method for stochastic optimization[C]//Proceedings of the 3th International Conference on Learning Representations. San Diego: ICLR, 2015: 1-13.

作者简介



李康和 男,1997 年 1 月出生于广东省湛江市. 现为华南师范大学人工智能学院硕士研究生. 主要研究方向为欺诈检测及其应用.
E-mail: licomen2022@163.com

黄震华(通讯作者) 男,1980 年 9 月出生于福建省莆田市. 现为华南师范大学计算机学院和人工智能学院教授、博士生导师. 主要研究方向为图神经网络、欺诈检测和推荐系统.
E-mail: huangzhenhua@m.scnu.edu