DOI:10.20079/j.issn.1001-893x.210725002
引用格式:王正用,白桦,吴笛,等.一种电力线网络上防御潜在攻击的鲁棒性消息传递机制[J].电讯技术,2023,63(4):556-562.[WANG Z Y, BAI H, WU D, et al. A robust message passing mechanism against potential attacks in power line networks[J]. Telecommunication Engineering, 2023, 63(4):556-562.]

一种电力线网络上防御潜在攻击的鲁棒性消息传递机制*

王正用[1],白　桦[1],吴　笛[1],褚如旭[2],王韬樾[2],林　斌[2]
(1. 浙江华云电力工程设计咨询有限公司,杭州 310004;2. 杭州智微易联电力科技有限公司,杭州 310030)

摘　要:网络结构上的信息传递(Message Passing,MP)机制是理解大型网络结构,实现网络节点属性预测的重要工具。然而,在电力线通信网络中,无法在终端和链路上部署复杂的身份识别验证的计算模块,导致无法建立可信的电力线通信网络 MP 机制,从而无法杜绝潜在攻击者的恶意对抗攻击行为,例如恶意更改电表数据、恶意隐藏终端故障等。为此,以电力节点匿名为动机,将电力节点位置隐藏至高斯噪声中,从而使攻击者无法定位具体的目标节点,抵制潜在的攻击行为。首先证明了节点定位定理,给出了图信息在傅里叶域中传递的数学模型,从而展示主流 MP 模型(图神经网络)定位节点的具体量化指标;然后设计了一个从高斯噪声中生成节点位置的生成器,实现节点匿名;最后设计了一个对抗生成网络,用来实现精确的电力线通信网络建模。实验表明,即使完全隐藏电力线中节点的位置,通过端到端训练,依然可以实现精确的 MP,从而构建更加可信的电力线通信网络理解机制。

关键词:电力线通信网络;消息传递;潜在攻击;恶意攻击;对抗攻击;对抗生成网络

开放科学(资源服务)标识码(OSID):

中图分类号:TN915.853　　**文献标志码**:A　　**文章编号**:1001-893X(2023)04-0556-07

A Robust Message Passing Mechanism against Potential Attacks in Power Line Networks

WANG Zhengyong[1], BAI Hua[1], WU Di[1], CHU Ruxu[2], WANG Taoyue[2], LIN Bin[2]
(1. Zhejiang Huayun Power Engineering Design Consulting Co., Ltd., Hangzhou 310004, China;
2. Hangzhou Zhiwei Yilian Power Technology Co., Ltd., Hangzhou 310030, China)

Abstract: Message passing(MP) mechanism on network structure is an important tool to understand large-scale network structure and realize attribute prediction of network nodes. However, in the power line communication(PLC) network, it is impossible to deploy complex identification and verification computing modules on the terminal and the link, resulting in the failure to build a trusted MP mechanism in the PLC network, so that the malicious counter attack behavior for potential attackers, such as malicious change of meter data, malicious concealment of terminal faults, etc., can not be eliminated. For this reason, this paper takes the power node anonymity as the motivation to hide the power node location in Gaussian noise, so that the attacker can not locate the specific target node and resist potential attacks. Firstly, the node localization theorem is proved, and the mathematical model of graph information transfer in the Fourier domain is given, so as to show the specific quantitative index of the mainstream MP model(graph neural network). Then, a generator is designed to generate node positions from Gaussian noise to realize node anonymity. Finally, a countermeasure generation network is designed to realize accurate modeling of PLC network. Experiments show that even if the location of nodes in the power line is completely hidden, accurate MP can still be achieved through end-to-end training, so as to build a more reliable PLC network understanding mechanism.

Key words: power line communication network; message passing; potential attack; malicious attack; adversarial attack; countermeasure generation network

*　**收稿日期**:2021-07-25;**修回日期**:2021-11-29
通信作者:王韬樾

0　引　言

电力线通信网络是一种复杂的网络结构,理解其内部承载的海量异构数据可以为构建更加智能化的工业物联网环境提供有力的支持。借助机器学习等高效模型,可以实现迅速、准确、可预测的电力线通信网络数据建模和预测。然而,机器学习由于其内生安全问题,极易遭受对抗攻击,所以无法提供确保可信的决策支持,这种潜在的威胁由此蔓延至其众多应用场景之中。同时,由于电力线通信网络终端的动态性和广泛性,无法部署严格的身份认证与鉴别模块,因此,理解和建模大规模电力线通信网络面临着潜在的攻击威胁,攻击者可以通过修改某物联网终端之间的链接拓扑,恶意篡改目标节点的属性,例如通过修改电表、协调器等终端之间的路由拓扑(为了隐蔽性,不直接作用于目标节点),使恶意消息通过信息传递(Message Passing,MP)机制传递到目标节点,从而实现其恶意行为。

目前,针对这种基于修改网络拓扑结构实现恶意行为的攻击,主要的防御手段来自以下两种方案:一是基于对抗训练的方案[1],通过对抗训练[2](Adversarial Training)建立能容纳所有潜在篡改模式的大容积模型,认知所有的潜在威胁,从而部署响应的防御措施;二是基于鲁棒聚合的方案[3],在构建更加鲁棒的聚合函数[4],进一步使模型能够识别和过滤潜在的扰动。然而,在假设所有边都是可扰动的前提下,潜在的攻击行为可能会对 MP 的鲁棒性带来很大的挑战,从而降低上述两种方案的有效性(后文将说明最先进的防御方案[5]会受到边缘扰动攻击,72.8%受保护的节点被错误分类为目标类别)。

本文的动机来自于对现有攻击手段的深刻观察,即在实际场景中,攻击者总是可以利用对电力线通信网络拓扑结构的边读取权限重新构建网络拓扑,从而误导 MP 机制。为此,本文提出了一个匿名图卷积神经网络(Anonymous Graph Convolutional Network,AN-GCN)来撤销 MP 模型的边读取权限,消除了潜在攻击者接触网络拓扑信息的机会,以及电力线网络中潜在的对抗攻击的可能。在 AN-GCN 中,电力线网络节点的位置是从噪声中随机产生的,同时保证了高精度的分类,即在保持节点位置匿名的同时分类节点。由于网络拓扑结构决定了节点的位置,因此位置的匿名性消除了 AN-GCN 读取边缘的必要性,从而最小化了修改边缘的可能性,有效解决了电力线网络上 MP 机制的脆弱性。

1　MP 模型的节点定位定理

由于 AN-GCN 具有生成节点位置的能力,本节提出一个节点定位定理作为 AN-GCN 的理论基础。以 GCN 作为 MP 机制的形式化模型,同时,因篇幅所限,本文所有的证明过程未给出,感兴趣的读者请微信扫描本文 OSID 码查看。

在 GCN 的训练阶段,信息不断地传递,这也可以看作是一种信号传输。因此,节点的特征是不断变化的,我们将其视为一个独立信号振动。本节建立了一个节点信号振动的数学模型来分析 GCN 如何分析给定的图形结构来定位节点。首先将每个节点特征的变化看作是信号随时间的振动,然后通过傅里叶变换将所有节点信号映射到同一正交基上。在此基础上,给出了 GCN 训练阶段的节点信号模型,直观地描述了 GCN 如何定位特定节点。建立节点信号模型的方法如图 1 所示。

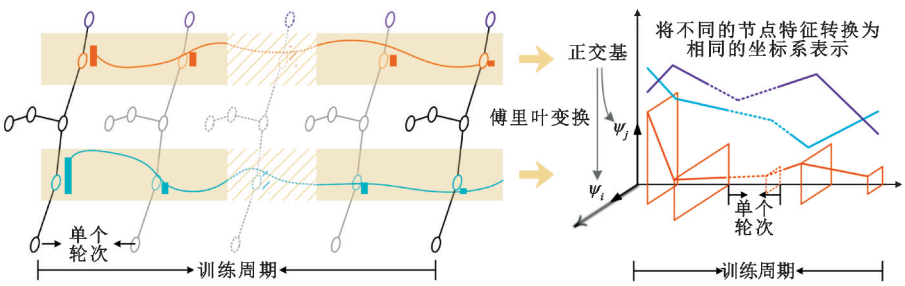

图 1　建立节点信号模型的方法

将节点特征随训轮次的变化视为独立的信号振动,通过傅里叶变换将所有信号变换为统一正交基下的表示。

符号表示 $\mathcal{G}=(f,\varepsilon)$ 表示一个网络结构,也称作图结构(Graph Structure),其中,f 是 N 个节点的节点属性集合,$f(i)$ 是节点 i 的属性,ε 网络中边的集合。在谱域图分析中一个潜在的算子是图的拉普拉斯算子,它定义为 $\boldsymbol{\Delta}=\boldsymbol{D}-\boldsymbol{A}$,其中度矩阵 $\boldsymbol{D}\sim\mathbb{R}^{N\times N}$,邻接矩阵 $\boldsymbol{A}\sim\mathbb{R}^{N\times N}$($\boldsymbol{D}$ 和 $\boldsymbol{A}$ 都是根据 ε 计算得来)。定义 $\mathcal{G}$ 的拉普拉斯矩阵是 $\boldsymbol{\Delta}$,同时 $\boldsymbol{\Delta}$ 的特征值是 $\lambda_1,\lambda_2,\cdots,$

λ_N,对应的特征矩阵是 $\boldsymbol{\Delta U}=\begin{pmatrix} u_1(1) & \cdots & u_N(1) \\ \vdots & & \vdots \\ u_1(N) & \cdots & u_N(N) \end{pmatrix}$,$u_l=(u_l(1),u_l(2),\cdots,u_l(N))^{\mathrm{T}}$ 是第 l 个特征向量,$u(l)=\{u_1(l),u_1(2),\cdots,u_N(l)\}$ 是由位置 l 处所有特征向量的值组成的行向量。为了方便,编号为 n 的节点被标记为节点 n。

1.1 节点定位定理

定理 1(节点定位定理) 给定一个要被 GCN 学习的图 $\mathcal{G}$,GCN 根据 $u(\alpha)$ 定位节点 α。

上述定理由以下两节证明。

1.2 节点信号模型

首先,给出特定频率下单个节点信号的傅里叶域坐标。

引理 1 给定频率 ν,节点 α 的傅里叶域中的信号坐标为

$$\hat{f}_\alpha[\nu]=\sum_{t=0}^{E-1}\bar{\theta}_t\Big(\sum_{i=1}^{N}c_i\mathrm{e}^{\lambda_i t}u_i(\alpha)\Big)\mathrm{e}^{-\mathrm{j}\frac{2\pi}{E}t\nu}。\tag{1}$$

式中:$\bar{\theta}_t$ 和 c_i 为常数;E 是总的训练轮次。

然后,对于节点 α,在傅里叶域中积分所有频率的信号以获得正交基表示的原始信号,在下面的引理中陈述。

引理 2 在训练轮次 t 时,α 的信号为

$$f_\alpha[t]=\sum_{\nu=0}^{E-1}\frac{2}{E}\left|\hat{f}_\alpha[\nu]\right|\cos[(2\pi\nu/E\epsilon)t\epsilon+\arg(\hat{f}_\alpha[\nu])]\mathrm{e}^{\mathrm{j}\frac{2\pi}{E}t\nu}。\tag{2}$$

式中:$\arg(\cdot)$ 是复数的辐角;ϵ 表示训练轮次的间隔,这里将其作为最小值,即 $\epsilon\to 0$。

1.3 初始状态和训练状态的节点信号

引理 3 节点 α 的初始信号为

$$f_\alpha[0]=2\bar{\theta}_0\mathrm{Sum}[u(\alpha)]。\tag{3}$$

接下来激活 $\mathcal{G}$ 上的 MP。由于 $\bar{\theta}$ 代表了常数,式(1)可以通过下式计算:

$$\hat{f}_\alpha[\nu]=\sum_{t=0}^{E-1}\Big(\sum_{i=1}^{N}\bar{c}_i\mathrm{e}^{\lambda_i t}u_i(\alpha)\Big)\mathrm{e}^{-\mathrm{j}\frac{2\pi}{E}t\nu}。\tag{4}$$

式中:$\bar{c}_i=\bar{\theta}_t c_i$。将式(4)代入式(2),节点信号的最终表示为

$$f_\alpha[t]=\lim_{\epsilon\to 0}\Big\{\sum_{\nu=0}^{E-1}\frac{2}{E}\Big|\sum_{t=0}^{E-1}\Big(\sum_{i=1}^{N}\bar{c}_i\mathrm{e}^{\lambda_i t}\underbrace{u_i(\alpha)}_{\text{from }\varepsilon}\Big)\mathrm{e}^{-\mathrm{j}\frac{2\pi}{E}t\nu}\Big|\cdot \cos[(2\pi\nu/E\epsilon)t\epsilon+\arg(\hat{f}_\alpha[\nu])]\mathrm{e}^{\mathrm{j}\frac{2\pi}{E}t\nu}\Big\}。\tag{5}$$

上式给出了统一坐标系下各节点的信号模型。通过观察,对于节点 α,在所有与 $\mathcal{G}$ 相关的初始条件中,只有 $u(\alpha)$ 出现在节点信号变化的过程中。换言之,在端到端训练带来的弹性系统中,只有 $u(\alpha)$ 是可控因子,因此将上述方程简化为

$$f_\alpha[t]=\mathcal{F}[u(\alpha)]。\tag{6}$$

节点 α 在端到端训练下的信号振动为

$$\begin{cases}\text{Initial}:2\bar{\theta}_0\mathrm{Sum}[u(\alpha)]\\ \text{Training}:\mathcal{F}[u(\alpha)]\end{cases}。$$

因此,GCN 驱动的节点 α 的特征振动可以被量化,并且在消息传递时它总是包含一个固定的因子 $u(\alpha)$。由于 GCN 根据输入 Δ 对训练阶段传递的信息进行量化,GCN 将 Δ 分解为拉普拉斯矩阵 $\boldsymbol{U}$,进一步将节点 α 的特征振动量化为 $u(\alpha)$,即 GCN 根据 $u(\alpha)$ 定位节点 α。

2 匿名图神经网络

根据定理 1,将拉普拉斯矩阵的每一行表示为一个独立的生成目标。本文使用谱图卷积[6](没有任何约束和简化规则)作为基本模型,并将其作为编码器,使用一个额外的完全连接的神经网络解码器,也就是说,基本的前向传播是

$$f^e=\sigma(\boldsymbol{U}g_\theta(\Lambda)\boldsymbol{U}^{\mathrm{T}}f),\tag{7a}$$

$$y_{\text{out}}=f^e\boldsymbol{W}^D。\tag{7b}$$

式中:f^e 是嵌入空间中的特征。在式(7a)中,$\boldsymbol{U}$ 包含图中边的信息,用高斯噪声产生的矩阵来代替它。式(7)是基本的信号前向传播。接下来,将详细说明如何改进式(7),从而使基本模型适应生成的节点位置。

2.1 从噪声中生成节点位置

如果 $u(n)$ 完全由噪声产生,则特定点将保持匿名,因此,将 $u(n)$ 作为生成目标。生成器的输出表示为 $u^G(n)$,它试图近似于底层的真实节点位置分布 $u(n)$。

接下来,定义输入噪声的概率密度函数(Probability Density Function,PDF)。为了使生成器能够定位特定的点,生成器的输入噪声受到目标点位置的约束。因此将输入噪声定义为具有一个与错峰高斯分布相同的 PDF,目的是保证噪声不仅满足高斯分布,而且彼此不重合,从而使产生的噪声在数轴上有序分布。

引理 4 给出最小概率 ε,N 个以 $x=0$ 为中心的高斯分布满足

$$P(x,n)\sim\mathrm{Norm}(2\partial(2n-N-1)\sqrt{\ln(\sqrt{2\pi}\partial\varepsilon)},\partial^2),$$

使每个分布的概率密度函数大于 ε,其中,Norm 是高斯分布,n 是节点编号,∂ 是标准差,ε 是设置的最小概率。

从错峰高斯噪声 $Z_v\sim P(x,v)$ 中生成 $u^G(v)$ 表示为 $u^G(v)=G(Z_v)$,从 G 中生成的 $\boldsymbol{U}$ 表示为 $\boldsymbol{U}^G$,该过程如图 2 所示。

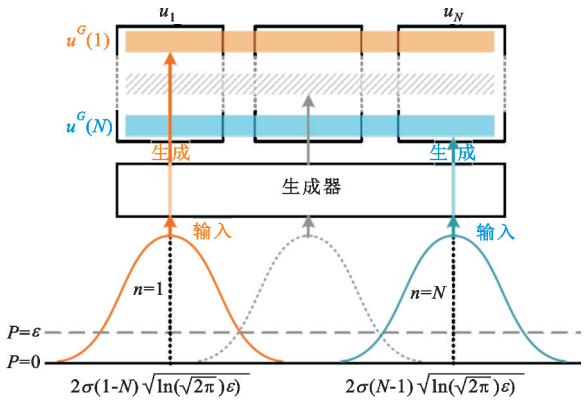

图 2　从错峰高斯噪声中生成的噪声作为生成器的输入

2.2　检测生成的节点位置

如图 3 所示,在给出生成 $u^G(n)$ 之后,设置一个判别器 D 来评价 $u^G(n)$ 的质量,并设计一个针对 G 和 D 的博弈规则,确保了 AN-GCN 可以通过 $u^G(n)$ 准确地对节点进行分类。使用分类质量作为评估指标来推动整个对抗训练,确保了 D 不仅能够区分由 G(非恶意,用于匿名化节点)生成的对抗性样本,而且能够提供准确的分类。

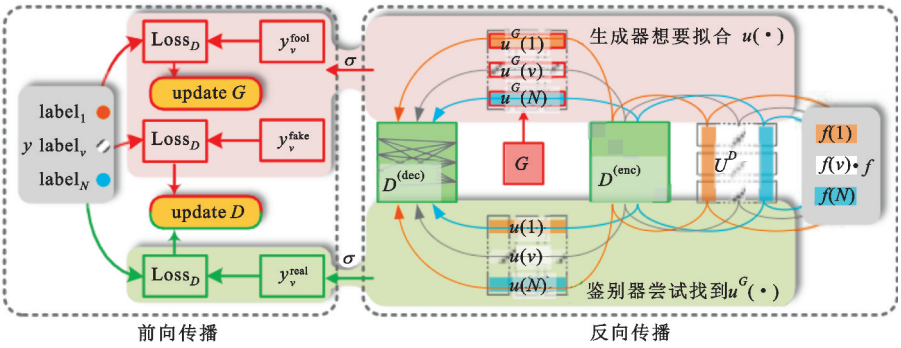

图 3　AN-GCN 示意图

具体来说,D 分为两部分:一部分是用于编码的带有可训练参数 D_{enc} 的对角矩阵;另一部分是用于解码的参数矩阵 D_{dec}。所以,基础模型的前向传播是

$$Y=\sigma(\boldsymbol{U}D_{\text{enc}}\boldsymbol{U}^{\text{T}}fD_{\text{dec}})。\tag{8}$$

考虑到编码器中有 $\boldsymbol{U}$ 和 $\boldsymbol{U}^{\text{T}}$,将两者标记为不同的矩阵:

$$Y=\sigma(\boldsymbol{U}^GD_{\text{enc}}\boldsymbol{U}^DfD_{\text{dec}})。\tag{9}$$

为了消除 AN-GCN 中节点的位置信息,通过独立生成不同的行来生成相应的 $u(\cdot)$ 组成 $\boldsymbol{U}^G$,右边的 $\boldsymbol{U}^D$ 根据 $\boldsymbol{U}^G$ 进行线性的拟合变化。因为 $\boldsymbol{U}^D$ 和 $\boldsymbol{U}^G$ 在动态变化,使用 $\boldsymbol{U}^{G,e}$ 和 $\boldsymbol{U}^{D,e}$ 表示两者在第 e 轮次的值。当生成 $\boldsymbol{U}^G$ 的第 l 行时,$\boldsymbol{U}^D$ 对应的列标记为 u_l^D,同时向 $u^G(l)$ 线性拟合。具体地说,其在训练轮次 e 中的值的通称公式是

$$u_l^{D,e}=\begin{cases}u_l,\text{s. t. } e=1\\ u_l^{D,e-1}+q(u^{G,e}(l)-u_l^{D,e-1}),\text{s. t. } e>1\end{cases}。$$

式中:$u^{G,e}(l)$ 是第 e 轮次的生成项;q 是自定义的线性拟合因子。因此,$\boldsymbol{U}^{D,e}$ 在训练阶段逐渐接近生成矩阵 $\boldsymbol{U}^{G,e}$。随着训练轮次的增加,式(8)中的 $\boldsymbol{U}$ 和 $\boldsymbol{U}^{\text{T}}$ 与 G 的关系如图 4 所示。

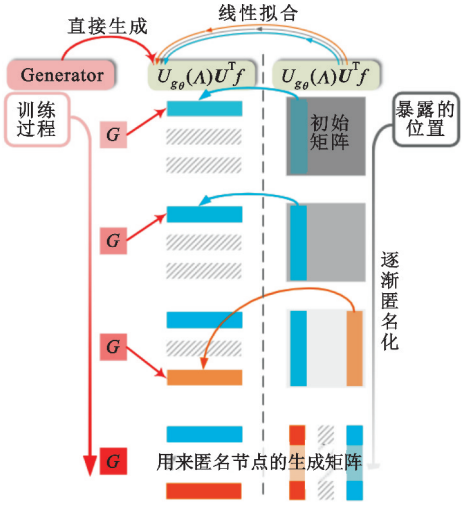

图 4　$\boldsymbol{U}$ 和 $\boldsymbol{U}^T$ 与 G 的关系

因此,AN-GCN 的前向传播如图 5 所示。

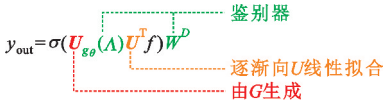

图 5　AN-GCN 的前向传播

目前,已经消除了存储在 AN-GCN 中的边缘信息,接下来将介绍允许节点匿名参与的训练方法。同时,为了保证对抗训练双方都有明确的目标,进一

步给出了生成器和鉴别器的优化方法。

2.3 AN-GCN 的前向传播

使用 $f^{e,G}(v)$ 和 $f^{e,D}(v)$ 表示节点 v 在使用 $u^G(v)$ 和 $u(v)$ 定位时的嵌入,在每一个训练轮次 e,首先计算节点 v 对应的 $u^G(v)$ 和 $U^{D,e}$,即从错峰高斯噪声中生成 $u^G(v)$,然后将 $\boldsymbol{U}^{D,e-1}$ 更新为 $\boldsymbol{U}^{D,e}$。接着,使用 D_{enc} 和 D_{dec} 评估本轮生成器的质量,通过获取节点嵌入 $f^{e,G}(v)$,使用 D_{dec} 解码至 v 的软标签(soft label)y_v,通过式(7a)得到

$$y_v=\boldsymbol{\sigma}[u(v)D_{\text{enc}}(\boldsymbol{\Lambda})\boldsymbol{U}^{\mathrm{T}}fD_{\text{dec}}]。\tag{10}$$

2.4 生成器的优化

鉴别器优化的目的是降低发生器的分类精度,同时提高鉴别器的分类精度,利用软标签对分类精度进行量化。对于拥有生成位置的节点 v,它的软标签被标记为 y_v^{fake}。y_v^{fake} 通过 $u^G(v)$、$\boldsymbol{U}^D$ 和鉴别器共同计算。针对训练鉴别器检测节点 v 产生位置的问题,由非防御类的 GCN 计算出真正的软标签 y_v^{real}。所以,y_v^{fake} 和 y_v^{real} 的计算方法为

$$y_v^{\text{fake}}=\boldsymbol{\sigma}[u^G(v)D_{\text{enc}}\boldsymbol{U}^Df(v)D_{\text{dec}}],\tag{11a}$$

$$y_v^{\text{real}}=\boldsymbol{\sigma}[u(v)D_{\text{enc}}\boldsymbol{U}f(v)D_{\text{dec}}]。\tag{11b}$$

为了检测具有生成位置的节点,鉴别器不仅要将真实节点分类为正确的类别,还要将虚假节点分类为其他随机类别,这在鉴别器的训练阶段降低了生成的性能。生成器和鉴别器的性能由损耗函数量化,其设计如下:

$$\text{Loss}_D(y)=\begin{cases}C[S(y),\text{label}_v],\text{s.t. } y=y_v^{\text{real}}\\ C[S(y),\text{RD}(\{\text{label}_\gamma\mid\gamma\neq v\})],\text{s.t. } y=y_v^{\text{fake}}。\end{cases}\tag{12}$$

式中:label_i 表示节点 i 的正确类别(one-hot 形式);$C(\cdot)$表交叉熵(示 Cross entropy)函数;$S(\cdot)$表示 Sigmoid 函数;RD$(\cdot)$表示随机抽样函数。之后,根据 $\text{Loss}_D(y)$,D 的梯度的计算方法为

$$\nabla_{\text{update}}^D=\nabla_{g_\theta^{D,\text{dec}}(\Lambda),\theta^{D,\text{enc}}}[\text{Loss}_D(y_v^{\text{real}})+\text{Loss}_D(y_v^{\text{fake}})]。\tag{13}$$

式中:$g_\theta^{D,\text{dec}}(\Lambda)$ 和 $\theta^{D,\text{enc}}$ 是 D_{enc} 和 D_{dec} 的可训练参数。最后,根据 ∇_{update}^D 更新生成器的参数。

2.5 鉴别器的优化

接下来,优化鉴别器用来欺骗训练良好的生成器,确保了 $u^G(v)$ 可以提供准确的节点分类。对于节点 v,在重新采样造成 Z_v 后,用来欺骗鉴别器的软标签计算方法如下:

$$y_v^{\text{fool}}=\boldsymbol{\sigma}[G(Z_v)D_{\text{enc}}\boldsymbol{U}^DfD_{\text{dec}}]。\tag{14}$$

在生成阶段,生成器尝试将节点 v 分类为正确的标签,因此 D 的损失函数是

$$\text{Loss}_D(y)=C[S(y),\text{label}_v],\text{s.t. } y=y_v^{\text{fool}}。\tag{15}$$

然后,为了保证 G 的输出能够被划分为正确的类别,D 的梯度计算方法为

$$\nabla_{\text{update}}^G=\nabla_{\theta^G}[\text{Loss}_D(y_v^{\text{fool}})]。\tag{16}$$

之后,根据 ∇_{update}^G 更新生成器的权重。注意,D 在生成器的训练阶段被冻结。AN-GCN 的示意图见图 3,算法伪代码如下:

Require:D 的权重 θ^D,包括用来编码的 $g_\theta^{D,\text{enc}}(\Lambda)$ 和用来解码的 $\theta^{D,\text{dec}}$;G 的权重 θ^G;D 和 G 的训练周期。

初始化 $\boldsymbol{U}^D=\boldsymbol{U}$
for D 的训练周期 do
  随机选取一个节点 v
  从 $P(z,v)$ 中获取噪声 Z_v
  $u^G(v)\leftarrow G(Z_v^D)$//生成
  $\boldsymbol{U}_v^D=U_v^D+q(u^G(v)-U_v^D)$//线性拟合
  计算 v 的假标签(式(11a))
  计算 v 的真标签(式(11b))
  计算∇_{update}^D(式(13))
  $\theta^D\leftarrow\theta^D-\nabla_{\text{update}}^D$
  for G 的训练周期 do
    从 $P(z,v)$ 中获取噪声 Z_v
    计算 y_v^{fool}(式(14))
    计算∇_{update}^G(式(16))
    $\theta^G\leftarrow\theta^G-\nabla_{\text{update}}^G$
  end for
end for
Ensure 训练好的生成器权重 θ^G

2.6 安全性分析

在训练阶段,随着 $\boldsymbol{U}^D$ 逐渐线性逼近 $\boldsymbol{U}$,AN-GCN 的前向传播为

$$Y=\begin{bmatrix}G(Z_1)\\ \vdots\\ G(Z_N)\end{bmatrix}D_{\text{enc}}[G(Z_1)^{\mathrm{T}},\cdots,G(Z_N)^{\mathrm{T}}]fD_{\text{dec}},\tag{17}$$

可简化为

$$Y=\text{AN-GCN}(\mathbb{N};g_\theta^{D,\text{dec}}(\Lambda),\theta^{D,\text{enc}})。\tag{18}$$

式中:$\mathbb{N}=\{1,2,\cdots,N\}$ 代表节点编号集合。相较于现有的非防御 GCN,

$$Y=\text{GCN}(f;A,\theta^E)。\tag{19}$$

AN-GCN 中节点的位置是不可见的,即 AN-GCN 可以匿名化节点,从而杜绝潜在的攻击。

3 实验结果

3.1 节点定位定理的数值实验

由于定理 1 是 AN-GCN 的理论基础,本部分通过数值实验验证定理 1 的正确性。设计了两组实验

来验证定理 1。在第一组实验中,观察了节点在嵌入空间中的位置偏差,该偏差是通过一个手动因子干扰相应的 $u(\cdot)$。如果在干扰 $u(v)$时,节点 v 的嵌入偏差明显大于其他节点的嵌入偏差,则证明$u(v)$对 v 的位置高度敏感。在第二组实验中,观察了删除节点后 $u(\cdot)$邻居的变化。如果对应的 $u(\cdot)$的变化随邻居顺序(与被删除节点的距离)的增加而不显著,则进一步证明 v 的具体位置 $u(v)$的值密切相关。两组实验共同证明了定理 1 的理论正确性。

3.1.1　$u(\cdot)$对节点嵌入的影响

首先观察节点 v 在轻微的扰动对应的 $u(\cdot)$后嵌入空间中的偏差。用 $\mathrm{Nbor}_{1^{\mathrm{th}}}(v,c_v)=\{v_1(1),v_1(2),\cdots,v_1(c_v)\}$表示 v 的前 c_v 个一阶邻居。其中,c_v 是主观地给出的,用来过滤阶数较低的邻居。给定一个干扰因子 δ,先后扰动$\mathrm{Nbor}_{1^{\mathrm{th}}}(v,c_v)$,所以,在每一个扰动周期内,给定一个扰动目标节点 v_i^{close},未被扰动的 $\boldsymbol{U}$ 更新为一个扰动矩阵 $\hat{\boldsymbol{U}}^{\delta,c_v}$,其中每的一行满足

$$\hat{u}^{\delta,c_v}(i)=\begin{cases}u(i),\text{s.t. } i=v_1(i)\\ \delta u(i),\text{s.t. } i\neq v_1(i)\end{cases}。\tag{20}$$

然后通过 $\hat{\boldsymbol{U}}^{\delta,c_v}$ 和一个训练良好的编码器将 f 嵌入到嵌入空间中。

之后,计算 $f^{e,\delta}$ 和 f^e 的欧氏距离[7](根据式(7a)),也就是说,$d_v^\delta=\|f^{e,\delta}-f^e\|_2$。在更变 δ 的同时,使用基于切比雪夫多项式的卷积核[6],在真实网络数据 Cora 数据集上进行实验。使 $c_v=16,\delta=1-\frac{k}{100},k\in\{1,2,\cdots,50\}$,结果如图 6 所示,使用 $f^e(v)$和 $f^{e,\delta}(v)$的欧氏距离度量嵌入精度。当 δ 扰动到 $\boldsymbol{U}$ 的第 v^{th} 行时嵌入精度迅速下降,同时在扰动其他位置时保持平稳。可以看到 $u(v)$对节点 v 的影响显著。

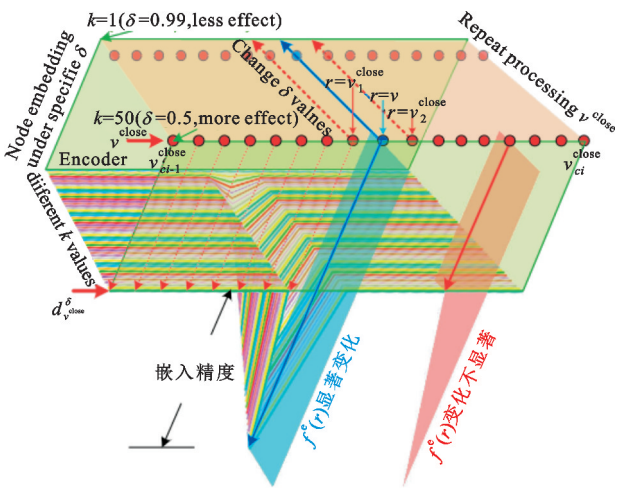

图 6　节点 v 最高度的前 16 个邻居接受扰动之后的偏差

3.1.2　节点位置改变时对 $u(\cdot)$的影响

通过删除 $\mathcal{G}$ 中的一个节点 τ 从而获得 $\mathcal{G}_\tau^{(d)}$,计算拉普拉斯矩阵 $\boldsymbol{L}_\tau^{(d)}\in\mathbb{R}^{(N-1)\times(N-1)}$和其特征矩阵 $\boldsymbol{U}_\tau^{(d)}=\{u^{(d)}(1),\cdots,u^{(d)}(N-1)\}\in\mathbb{R}^{(N-1)\times(N-1)}$。为了保持 $\mathcal{G}_\tau^{(d)}$ 连接良好,与 τ 连接的所有边将被遍历地重新连接。具体地,规定 ω_{ij} 和 $\omega_{ij}^{(d)}$ 分别表示 $\mathcal{G}$ 和 $\mathcal{G}_\tau^{(d)}$ 中节点 i 和 j 之间边的权重,$\mathcal{G}_\tau^{(d)}$ 中所有边的权重计算方法为

$$\omega_{ij}^{(d)}=\begin{cases}\omega_{ij},\omega_{\tau i}=0\text{ or }\omega_{\tau j}=0\\ \omega_{ij}+\dfrac{\omega_{\tau i}+\omega_{\tau j}}{2},\omega_{\tau i}\neq 0,\omega_{\tau j}\neq 0\end{cases}。\tag{21}$$

在获取 $\mathcal{G}_\tau^{(d)}$ 后,重新计算对应的 $U_\tau^{(d)}$,重新读取 τ 的 β^{th} 阶邻居$\mathrm{Nbor}_{\beta^{\mathrm{th}}}(\tau,\cdot)=\{\tau_\beta(1),\tau_\beta(2)\cdots\}$。因此,获取到 $\mathcal{G}_\tau^{(d)}$ 中$\mathrm{Nbor}_{\beta^{\mathrm{th}}}(\tau,\cdot)$所对应的集合 $u^{(d)}(\tau_\beta(\cdot))=\{u(\tau_\beta(1)),\cdots,u(\tau_\beta(N-1))\}$。类似地,$u(\tau_\beta(\cdot))$表示从 $\mathcal{G}$ 获取的集合,两者之间变化的定量表示为

$$C[u(\tau_\beta(\cdot)),u^{(d)}(\tau_\beta(\cdot))]=\sum_{i=1}^{N-1}[\ln|u_i(\tau_\beta(\cdot))|^2-\ln|u_i^{(d)}(\tau_\beta(\cdot))|^2]。\tag{22}$$

通过检查不同的 β 并计算 $C(\cdot)$,结果如图 7 所示。根据从大到小的连接数来选择前 500 个节点。在删除一个节点 τ 后,τ 的 β 阶邻居的 $u^{(d)}(\tau_\beta(\cdot))$的变化情况如图 7 中右图所示。一阶邻居的 $u^{(d)}(\tau_\beta(\cdot))$变化最大(纵轴是$-C$),也就是说,在 τ 被删除后,一阶邻居 $u(\tau_1(\cdot))$对应的 $u(\cdot)$变化最显著,同时 $u(\tau_2(\cdot))$和 $u(\tau_3(\cdot))$变化逐渐降低。由于删除节点 τ 后将显著影响其一阶邻居,同时随着阶数增加而降低,证明了节点 τ 的位置与$u(\tau)$是不可分的。

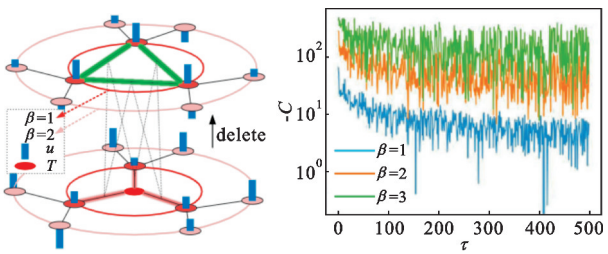

图 7　删除节点后不同位置 $u(\cdot)$的变化情况

3.2　AN-GCN 的有效性

由于 AN-GCN 消除了潜在扰动的可能性,因此主要对 AN-GCN 的精度进行评估,并将其对训练集扰动的鲁棒性作为次要评估项目。由于 AN-GCN

的应用场景是用户能够保证 AN-GCN 在干净图上训练,因此中毒攻击场景不是本文的主要问题,但本文仍然证明了 AN-GCN 对训练集中的扰动具有鲁棒性。

首先评估 AN-GCN 在 Cora 数据集上的准确性。本文希望 AN-GCN 在从噪声中产生位置的同时能够准确地对节点进行分类,从而用分类精度来评价 AN-GCN 的有效性。由于生成器不直接生成节点嵌入,而是通过与鉴别器的协作来预测节点,因此通过 G 生成的位置来表示,acc_G 为分类精度,acc_D 为鉴别器的分类精度。此外,使用 $\mathrm{acc}_{\mathrm{GCN}}$ 来表示单层 GCN(卷积核采用对称规范化拉普拉斯矩阵[8])的精度,并与相同的 D 作比较。此外,为了更直观地展示 AN-GCN 的优势,在训练阶段将嵌入空间可视化,结果如图 8 所示。可以看出,acc_D 和 acc_G 同时上涨。当训练轮次大于 1 400 时,acc_G 保持了较高切平稳的状态。选择 1 475 轮次的 G 作为最终的模型。节点分类的精确度为 0.822 7,$\mathrm{acc}_{\mathrm{GCN}}$ 的最高值为 0.793 4。实验结果表明,在相同的卷积核设计下,AN-GCN 不仅有效地保持了节点的匿名性,而且分类精度比现有的非防御 GCN 提高了 0.029 3。

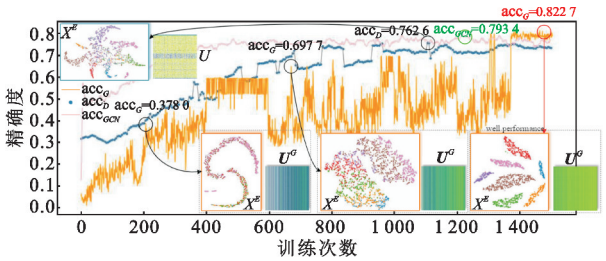

图 8 训练阶段的分类精度和节点嵌入的可视化

4 结 论

本文针对电力线网络中 MP 机制所遭受的潜在攻击威胁,提出了 AN-GCN 来防御边缘扰动攻击。具体地,提出并证明了一个节点定位定理,设计了一个基于交错高斯噪声的节点位置发生器来匿名化节点的位置,以及一个基于谱图卷积的鉴别器来保证高精度的分类。最后,给出了所设计的发生器和鉴别器的优化方法。安全性分析表明,AN-GCN 在防御边缘扰动攻击方面是安全的。广泛的评估验证了一般边缘扰动攻击模型的有效性,并证明了 AN-GCN 在节点分类任务中具有比非防御 GCN 更高的精度。

参考文献:

[1] DAI Q Y,SHEN X,ZHANG L,et al. Adversarial training methods for network embedding[C]//Proceedings of 28th International Conference on World Wide Web. San Francisco:IEEE,2019:329-339.

[2] WANG Y S, ZOU D F, YI J F, et al. Improving adversarial robustness requires revisiting misclassified examples [C]//Proceedings of 8th International Conference on Learning Representation. Addis Ababa: IEEE,2020:1-7.

[3] GEISLER S,ZÜGNER D,GÜNNEMANN S. Reliable graph neural networks via robust aggregation[C]//Proceedings of 34th Conference on Neural Information Processing Systems. Vancouver:ACM,2020:13272-13284.

[4] HAMILTON W L, YING R, LESKOVEC J. Inductive representation learning on large graphs [C]// Proceedings of 31st International Conference on Neural Information Processing System. Long Beach:ACM,2017: 1025-1035.

[5] LI Y,JIN W,XU H,et al. DeepRobust:a platform for adversarial attacks and defenses[C]//Proceedings of 35th AAAI Conference on Artifical Intelligence.[S.l.]: AAAI,2021:16078-16080.

[6] DEFFERRARD M, BRESSON X, VANDERGHEYNST P. Convolutional neural networks on graphs with fast localized spectral filtering[C]//Proceedings of 30 th Advance Neural Information Processing System. Barcelona:IEEE,2016:3844-3852.

[7] WANG L,ZHANG Y,FENG J. On the Euclidean distance of images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2005,27(8):1334-1339.

[8] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[C]//Proceedings of 5th International Conference on Learning Representation. San Juan:IEEE,2016:1-9.

作者简介:

王正用 男,1988 年生于浙江温州,中级工程师,主要研究方向为配电网规划、配网工程设计。

白 桦 男,1980 年生于北京,硕士,高级工程师,主要研究方向为电网规划研究、工程设计。

吴 笛 男,1983 年生于浙江嘉兴,硕士,工程师,主要研究方向为新能源规划研究、工程设计。

褚如旭 男,1989 年生于浙江台州,初级工程师,主要研究方向为软件开发。

王韬樾 男,1988 年生于陕西西安,初级工程师,主要研究方向为配电物联网、电力通信系统。

林 斌 男,1987 年生于浙江温州,初级工程师,主要研究方向为配电物联网。