软件学报 ISSN 1000-9825, CODEN RUXUEW
2026,37(4):1511−1530 [doi: 10.13328/j.cnki.jos.007524] [CSTR: 32375.14.jos.007524]
©中国科学院软件研究所版权所有.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

面向欺诈检测的风险感知动态聚合图联邦学习*

杨家震[1], 邱 天[1], 陈可嘉[1], 段明江[1], 蒋 健[2], 胡泽远[2], 宋明黎[1], 冯尊磊[1]

[1](区块链与数据安全全国重点实验室 (浙江大学), 浙江 杭州 310027)
[2](中国移动 (浙江) 创新研究院有限公司, 浙江 杭州 310012)
通信作者: 冯尊磊, E-mail: zunleifeng@zju.edu.cn

摘 要: 随着信息技术的迅猛发展, 欺诈行为在金融交易、社交网络与评论系统等多个领域呈现出日益复杂化和多样化的趋势, 给传统欺诈检测技术带来了严峻挑战. 当前主流的基于图神经网络的方法虽然在单机构数据环境中表现出色, 但由于涉及用户敏感信息, 难以实现跨机构间的数据共享与协作, 进而限制了模型的训练效果与泛化性能. 联邦学习作为一种新兴的隐私保护分布式学习范式, 为跨机构协作训练提供了可行途径, 但现有图联邦学习方法多针对通用图任务设计, 难以适应欺诈检测中普遍存在的类别分布不平衡和数据异构性问题, 导致在欺诈样本识别方面表现不佳. 为应对上述挑战, 提出一种面向欺诈检测的风险感知动态聚合图联邦学习方法 (FedRPDA), 旨在有效应对跨机构的复杂欺诈风险事件识别. FedRPDA 包括两项关键策略: 典型风险动态聚合策略通过衡量客户端图中欺诈节点的结构性风险强度, 并结合具有时间衰减特性的动态权重映射机制来自适应地调整客户端的聚合权重, 从而在数据异构条件下增强全局模型对正常样本与典型欺诈样本的判别能力; 多样化风险平均聚合策略结合基于变分扰动的欺诈样本特征增强机制与全局原型引导的对比学习机制, 有效提升模型对结构多样、数量稀少的非典型欺诈样本的表征能力, 促进其在特征空间中向共性异常靠拢, 进一步提升模型在复杂欺诈风险场景下的识别鲁棒性. 在多个真实欺诈检测数据集上的实验结果表明, FedRPDA 在检测性能与训练收敛效率方面显著优于现有图联邦学习基线方法, 展现出良好的泛化能力与实际应用潜力.

关键词: 欺诈检测; 联邦学习; 图神经网络; 多样化风险
中图法分类号: TP18

中文引用格式: 杨家震, 邱天, 陈可嘉, 段明江, 蒋健, 胡泽远, 宋明黎, 冯尊磊. 面向欺诈检测的风险感知动态聚合图联邦学习. 软件学报, 2026, 37(4): 1511–1530. http://www.jos.org.cn/1000-9825/7524.htm

英文引用格式: Yang JZ, Qiu T, Chen KJ, Duan MJ, Jiang J, Hu ZY, Song ML, Feng ZL. Risk Perception Dynamic Aggregation Graph Federated Learning for Fraud Detection. Ruan Jian Xue Bao/Journal of Software, 2026, 37(4): 1511–1530 (in Chinese). http://www.jos.org.cn/1000-9825/7524.htm

Risk Perception Dynamic Aggregation Graph Federated Learning for Fraud Detection

YANG Jia-Zhen[1], QIU Tian[1], CHEN Ke-Jia[1], DUAN Ming-Jiang[1], JIANG Jian[2], HU Ze-Yuan[2], SONG Ming-Li[1], FENG Zun-Lei[1]

[1](State Key Laboratory of Blockchain and Data Security (Zhejiang University), Hangzhou 310027, China)
[2](China Mobile (Zhejiang) Research & Innovation Institute Co. Ltd., Hangzhou 310012, China)

Abstract: With the rapid development of information technology, fraudulent behaviors in multiple fields such as financial transactions, social networks, and review systems show an increasingly complex and diversified trend, which poses a serious challenge to traditional fraud detection techniques. Although current mainstream graph neural network-based methods perform well in single-agency data

* 基金项目: 国家重点研发计划 (2022YFB2703100)
本文由“高可信自适应机器学习”专题特约编辑王熙照教授、张敏灵教授、曹付元教授推荐.
收稿时间: 2025-04-28; 修改时间: 2025-06-30, 2025-08-15; 采用时间: 2025-08-20; jos 在线出版时间: 2025-09-02
CNKI 网络首发时间: 2026-01-02

environments, cross-agency data sharing and collaboration are difficult due to the involvement of sensitive user information, which in turn limits the training effectiveness and generalization performance of the model. Federated learning, as an emerging privacy-preserving distributed learning paradigm, provides a feasible way for cross-agency collaborative training, but existing graph federated learning methods are mostly designed for general graph tasks, making them difficult to adapt to the class imbalance and data heterogeneity problems prevalent in fraud detection, resulting in poor performance in fraud sample identification. To address the above challenges, this study proposes a risk perception dynamic aggregation graph federated learning method (FedRPDA) for fraud detection, aiming to effectively deal with complex fraud risk event recognition across organizations. FedRPDA includes two key strategies: the typical risk dynamic aggregation strategy measures the structural risk intensity of fraudulent nodes in the client graph and combines it with a dynamic weight mapping mechanism with temporal decay characteristics to adaptively adjust the aggregation weights of clients, thus enhancing the global model's ability to discriminate between normal samples and typical fraud samples under heterogeneous data conditions; the diversified risk average aggregation strategy integrates a variance perturbation-based feature enhancement mechanism for fraud samples with a global prototype-guided contrastive learning mechanism, which effectively improves the model's ability to represent structurally diverse and scarce a typical fraud samples, promotes their convergence toward common anomalies in the feature space, and further enhances the model's robustness in recognizing complex fraud risk scenarios. Experimental results on several real-world fraud detection datasets show that FedRPDA significantly outperforms existing graph federated learning baseline methods in terms of detection performance and training convergence efficiency, and demonstrates good generalization ability and practical application potential.

Key words: fraud detection; federated learning; graph neural network (GNN); diverse risk

数字信息产业的高速发展在推动社会智能化进程的同时, 也为不法分子提供了前所未有的作案手段. 尤其是伴随着互联网和人工智能技术的革新, 欺诈行为在金融交易[1,2]、社交网络[3,4]和评论系统[5,6]等多个领域呈现出专业化、规模化和复杂化的特点. 在金融交易场景中, 信用卡盗刷、洗钱及虚假交易等欺诈行为层出不穷, 其攻击手段隐蔽、行为迅速, 给银行和支付平台带来了严峻挑战, 甚至威胁金融体系的稳定. 社交网络中的欺诈者依托虚假身份散播误导性信息, 诱导用户上当受骗, 而虚假内容的辨别难度较大, 导致经济损失和社会风险加剧. 评论系统中的欺诈行为通过虚假评论、操控评分等手段误导消费者决策, 干扰市场公平竞争, 甚至形成大规模商业操纵. 这些欺诈模式普遍呈现出隐蔽性强、手段多样、规模化作案等特点, 给传统检测方法[7,8]带来了极大挑战. 为此, 学术界与工业界正积极探索新型检测机制, 以提升欺诈识别的鲁棒性和实时性, 为社会治理与风险防控提供坚实的技术保障.

近年来, 深度学习技术[9]凭借其强大的表征学习能力, 在高维非线性数据建模方面表现出色, 已被广泛应用于欺诈检测场景. 然而, 欺诈行为往往不仅依赖个体特征, 还涉及多个实体之间的复杂交互, 使得数据天然呈现出图结构属性, 因此图神经网络 (graph neural network, GNN)[10]被引入为欺诈检测提供了一种更具表现力的建模范式, 通过多层次信息聚合机制挖掘节点间潜在关联, 有效提升欺诈检测的精度和鲁棒性. 尽管基于 GNN 的欺诈检测方法取得了显著成效, 但该领域依然面临诸多现实挑战, 其中数据的安全性与隐私保护问题尤为突出. 欺诈检测通常涉及高度敏感的用户数据, 而受限于隐私保护法规及机构间的数据壁垒, 传统的集中式模型难以直接访问跨平台的全量数据, 导致模型泛化能力受限. 联邦学习[11]作为一种分布式机器学习范式, 能够在保护数据隐私的前提下实现跨机构协同训练, 有效解决数据孤岛问题. 为此, 研究者将联邦学习与图神经网络相结合, 提出图联邦学习[12], 使参与方能够基于本地的图数据进行模型训练, 通过联邦聚合机制实现模型参数共享与更新, 提升全局检测能力.

尽管图联邦学习在跨机构欺诈检测中展现出巨大潜力, 其在实际应用中仍面临两个关键挑战.

(1) 跨机构数据的高度异构性: 不同机构的图数据在类别分布、节点特征分布及拓扑结构等方面存在显著差异, 这种数据异构性不仅加剧了跨平台模型的训练难度, 也导致模型难以统一建模并在各场景中保持稳定的泛化性能.

(2) 欺诈数据的极端类别不平衡性: 欺诈检测任务中的正常活动数量远远超过异常活动, 导致训练过程中模型倾向于过度拟合多数类样本, 忽视少数类欺诈样本的特征表达, 这种偏向性使得模型难以有效捕捉异常模式, 在实际检测中引发较高的误报率.

数据异构性是导致联邦学习收敛速度下降和模型性能不佳的主要原因[13]. 非独立同分布 (Non-IID) 数据会引发局部模型参数更新的方向不一致, 导致在全局模型聚合过程中产生次优解, 当前主流优化方案主要聚焦于客户

端训练机制与服务端聚合策略两个方面进行改进. 在客户端训练阶段, 通过在损失函数中引入正则项约束模型更新方向来避免局部模型的发散[14−16]、预测图结构中缺失的邻居或链接来增强子图数据[17,18]以及采用多任务联合训练[19,20]等方式增强本地模型的质量. 在服务端聚合方面, 经典的 FedAvg 方法[11]依据客户端的样本数量进行加权聚合, 但在异构数据的现实场景中, 单纯依赖样本量的加权策略难以获得理想的模型性能, 因此提出了多种改进的聚合策略. 服务端可维护一组独立的测试数据, 通过 Shapley 值[21]、留一法[22]等方式评估各客户端模型的测试表现、利用客户端额外上传的梯度信息进行相似度计算[23−25]或根据图数据中的结构信息评估客户端的重要性[26,27], 据此动态调整聚合权重.

然而, 现有方法大多聚焦于优化联邦学习的训练稳定性与聚合策略, 虽然在一定程度上缓解了由数据异构性带来的影响, 但缺乏对欺诈风险分布特性的深入建模, 尤其是在面对极端类别失衡与异常模式多样化的场景下, 往往难以有效识别潜在的高风险样本. 此外, 当前的图联邦欺诈检测相关研究[28−31]多基于小规模、结构单一或类别相对均衡的数据集, 未能充分模拟真实跨机构欺诈检测环境的复杂性. 如图 1 所示, 在实际的欺诈检测场景中, 欺诈样本仅占据极小比例, 且由于机构间数据隔离以及欺诈行为的刻意隐蔽等因素, 欺诈类别内部存在高度的拓扑异质性 (如图 1 中的节点 1、2). 这种分布特性将导致现有联邦学习框架产生系统性认知偏差: 在全局模型聚合过程中, 参数更新方向受控于数量庞大的正常样本构成的分布主体以及具备典型拓扑特征 (如高频交互行为) 的欺诈样本, 而那些多样性较高但是数量稀少的非典型欺诈模式被边缘化, 显著削弱了模型在此类样本上的判别能力.

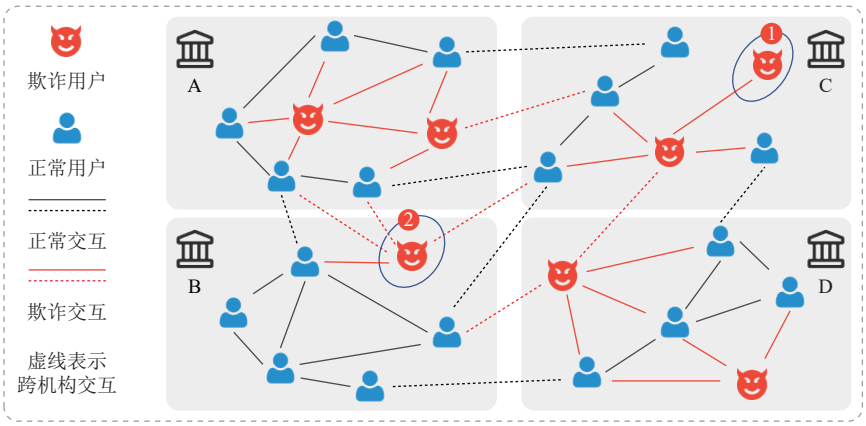


图 1　基于图结构的欺诈检测场景示例

为解决上述问题, 本文提出了一种面向欺诈检测的风险感知动态聚合图联邦学习方法 FedRPDA. 该方法在训练过程中引入多样性敏感适应机制, 旨在兼顾对典型欺诈模式的快速建模与对非典型欺诈模式的有效感知, 以更具针对性的方式增强图联邦学习方法在高风险场景下的建模能力. 核心设计包括两项关键策略: 一是典型风险动态聚合策略, 基于 PageRank 随机游走方法量化客户端本地图数据中欺诈节点的结构性风险强度, 计算其相对影响指数作为贡献度, 并通过时间衰减函数动态调整聚合权重, 使模型在训练初期优先聚焦于高风险、代表性强的客户端, 从而加速全局模型对典型欺诈样本的全局建模与泛化能力的提升; 二是多样化风险平均聚合策略, 在全局原型引导的对比学习阶段, 结合基于变分扰动的欺诈样本特征增强机制平衡两类数据, 促进多样化欺诈节点缩小类内方差的同时扩大与正常节点的类间距离, 引导模型构建更具判别性的表征空间, 从而提升其对结构多样、数量稀少的非典型欺诈样本的识别能力. 通过这两种策略的协同作用, 可以有效提升模型在高度不平衡和复杂风险环境中对多样化欺诈行为的检测能力.

综上所述, 本文的贡献主要包括以下 3 个方面.

(1) 提出了一种面向欺诈检测的风险感知动态聚合图联邦学习方法 FedRPDA, 在保护数据隐私的同时, 为跨机构场景下复杂欺诈风险的检测提供了新的解决思路.

(2) 设计典型风险动态聚合与多样化风险平均聚合两种策略, 使模型在训练过程中既能快速建模典型欺诈行

为, 又能兼顾对结构复杂、样本稀少的非典型欺诈行为的敏感性.

(3) 在多个真实欺诈场景下的图数据集上进行了大量实验验证, 结果表明所提出的 FedRPDA 方法在收敛效率、检测效果等方面均优于现有先进方法, 展现了其在实际应用中的有效性与广泛适用性.

本文第 1 节介绍图欺诈检测、联邦学习以及面向欺诈检测的图联邦学习的研究进展. 第 2 节给出图联邦学习的基本概念与相关定义. 第 3 节详细描述本文提出的风险感知动态聚合图联邦学习方法 FedRPDA. 第 4 节通过一系列实验验证所提出的 FedRPDA 在多种欺诈场景下的有效性与鲁棒性. 最后在第 5 节对全文进行总结归纳.

1 相关工作

1.1 图欺诈检测

在图欺诈检测领域, 近年来的研究主要集中在使用图神经网络来识别异常行为. GNN 通过消息传递机制实现节点之间的特征聚合和表示更新, 能够有效捕捉节点的属性信息和结构关联性, 实现对异常节点的识别. 现有的研究主要通过优化模型架构来增强模型的表示能力, 以及应对数据不平衡所带来的问题. 在优化模型架构方面, GUCNH[32]提出了一种融合门控机制的图卷积网络模块, 通过门控单元选择和组合模型在图卷积前后的特征, 从而得到更可靠的节点表示. GCCAD[33]利用图对比学习编码器, 通过无监督学习捕捉图结构特征进行异常检测, 取得了良好效果. GADAM[34]使用基于多层感知机的无冲突方式获取局部异常分数, 并采用基于混合注意力的自适应消息传递机制, 使节点能够选择性吸收周围正常或异常信号, 增强异常检测效果. 在处理数据不平衡问题方面, CAMD[35]提出了一种融合类别感知机制的不一致图神经网络框架, 结合半监督与自监督图对比学习方法, 有效提升了在数据不一致、类别不平衡与标签稀缺条件下的恶意节点检测性能. GDN[36]通过对异常节点和正常节点采取不同策略并约束关键异常特征, 显著缓解了图异常检测中的结构分布偏移. 尽管上述方法在图欺诈检测中取得了积极进展, 但大多聚焦于单一机构内部的图结构建模, 忽视了跨机构欺诈中多个机构间的交互. 例如, 金融欺诈团伙可能利用多个账户在不同机构间进行资金转移以掩盖其非法活动, 仅依赖局部图信息难以全面有效地处理这一问题.

1.2 联邦学习

联邦学习作为一种分布式机器学习范式允许多个客户端协作训练模型, 同时保证数据隐私. McMahan 等人[11]首次提出了联邦学习的概念及其应用场景, 并提出了 FedAvg 算法, 之后的研究都在此基础之上去解决联邦学习所面临的不同挑战. 现有的工作主要从两个角度开展研究: 一个是稳定本地的训练过程, 二是优化模型聚合策略. 为了应对客户端数据的非独立同分布问题, 研究者们提出了多种方法来稳定本地训练过程, FedProx[14]通过引入与全局模型距离的正则化项来限制本地模型参数, MOON[15]从表征的角度利用模型级别的对比学习来减少全局模型表示和本地模型表示的差异, FedProto[37]是第 1 个在异构联邦学习中使用原型聚合知识的框架, 通过使本地原型和全局标准保持一致来规范本地训练. 在模型聚合方面, Shapley 值[38]被认为是一个公平有效地衡量不同客户端模型在联邦学习任务中贡献的方法, 但其计算成本过高, 达到了 $O(N!)$ 的计算时间复杂度, 不适用于联邦学习的实际应用, 尽管一些方法[21,39]对其进行了优化, 但仍然需耗费大量时间, 而且该方法要求在服务端维护一个辅助测试集, 这在现实场景中往往不切实际.

图联邦学习作为联邦学习的一个特殊分支, 将其理念扩展到图结构数据上, 实现在保护隐私的同时联合多个机构中的图数据进行联合训练. 相比于传统的联邦学习, 它还需额外应对图数据的异构性和复杂性. FedSage+[17]设计了一个缺失邻居生成器, 用于处理局部子图之间可能缺失的边, 将受损子图修正为较为完整的子图, 实现本地图数据增强. FedGTA[40]利用图的拓扑结构和节点属性进行编码, 计算局部平滑置信度和邻居特征的混合矩衡量子图分布, 该方法仅聚合相似数据分布的客户端, 并赋予平滑图训练的模型更高权重. FGSSL[41]关注节点级语义和图级结构, 设计对比损失函数来处理客户端的局部失真问题, 对齐同类节点的局部和全局表示并拉远不同类别节点的距离, 以增强模型对节点的区分能力. FedGL[42]将预测结果和节点嵌入上传到服务器, 以导出全局伪标签和全局伪图, 利用这些全局知识缓解不同客户端的异构性. 然而, 现有图联邦学习方法主要面向节点分类、链接预测等通用

图任务, 其优化目标主要集中于缓解由数据异构性引发的训练不稳定问题. 受限于图数据本身的高度复杂性与多样性, 现有方法在应对不同类型图任务时泛化能力有限, 尤其在欺诈检测任务中, 难以有效建模欺诈风险的分布特征, 导致检测性能受限. 因此, 亟待设计一种更具针对性的联邦学习方法, 以更好地满足实际欺诈检测场景下的应用需求.

1.3 面向欺诈检测的图联邦学习

不同地区与机构间的数据隔离与隐私保护限制了跨平台欺诈数据的共享, 使得能够有效处理跨平台复杂欺诈关联关系的图联邦学习成为适合的解决方案. 文献 [28] 是首次将图联邦学习应用于反洗钱交易检测, 提出了结合图结构建模与联邦学习的协同平台. 2SFGL[29]提出了一种两阶段的联邦图学习方法, 首先通过多方图的虚拟融合构建增强图结构, 随后在虚拟图上进行模型训练与推理. FedGAT-DCNN[30]将图注意力网络与扩张卷积相结合, 提升模型对新型欺诈模式的适应能力. PPSED[31]结合客户端状态感知、梯度量化与 DDPG 策略优化, 实现了在数据异构与隐私约束条件下对公共安全事件的高效检测. 目前, 基于图联邦学习的欺诈检测研究仍处于起步阶段, 现有方法多基于小规模、结构相对简单的通用图数据集, 并采用理想化的数据划分策略进行实验验证, 尚未充分挖掘欺诈场景的独特数据特性, 在处理跨平台复杂关联结构和高度异构数据方面能力有限, 难以全面应对真实环境中的欺诈风险.

不同于已有研究, 本文在深入分析欺诈场景中数据特性的基础上, 提出了 FedRPDA 方法, 通过设计典型风险动态聚合与多样化风险平均聚合策略, 在保护数据隐私的前提下精准建模不同类型的欺诈风险, 为复杂场景下的跨机构欺诈检测提供了一种新颖而有效的解决方案. 此外, 在多个更贴近实际应用场景的大规模真实欺诈检测数据集上进行系统实验, 进一步验证了所提方法的有效性与实际应用价值.

2 预备知识

在一个联邦图学习系统中, 假设存在 N 个客户端与中央服务器 S 进行通信协作, 每个客户端 n 拥有自身的图数据 $G_n = (V_n, E_n, X_n, Y_n)$, $n \in \{1, 2, \ldots, N\}$, 其中 V_n、E_n、X_n 和 Y_n 分别表示图 G_n 的节点集合、边集合、节点特征集和节点真实标签集. 客户端的模型参数集合表示为 $\mathcal{W} = \{w_1, w_2, \ldots, w_N\}$, 服务端阶段性地收集客户端上传的模型参数, 通过聚合的方式来学习全局模型参数 w_g.

出于数据隐私的原因, 在整个训练过程中, 各个客户端是相互独立的, 不可以访问到其他客户端的图数据. 对于每个客户端, 定义其损失函数为 $\mathcal{L}_n$, 联邦图学习的全局优化目标可以表示为:

$$\min_w \mathcal{L}(w) = \sum_{n=1}^{N} \frac{|V_n|}{\sum_{n'=1}^{N} |V_{n'}|} \mathcal{L}_n(w_n) \tag{1}$$

其中, $|V_n|$ 表示第 n 个客户端图数据中的节点数量.

以 FedAvg[11]为例, 在正式开始训练之前, FedAvg 初始化模型参数 w^0 并广播至所有客户端, 在第 t 个通信轮次的训练过程中, 执行流程可以大致描述为以下步骤.

(1) 本地模型更新. 每个客户端 n 使用自身的图数据 G_n 训练模型, 并更新局部模型参数 $w_n^{t+1} \leftarrow w_n^t$.

(2) 参数上传. 客户端将更新后的局部模型参数 w_n^{t+1} 上传至中央服务器 S.

(3) 参数聚合. 服务端收集客户端上传的模型参数, 通过加权聚合的方式更新全局模型参数 $w_g^{t+1} \leftarrow \sum_{n=1}^{N} \frac{|V_n|}{\sum_{n'=1}^{N} |V_{n'}|} w_n^{t+1}$.

(4) 全局模型下发. 服务端将更新后的全局模型 w_g^{t+1} 广播给所有的客户端, 客户端使用此参数更新自己的局部模型.

3 风险感知动态聚合图联邦学习方法

在本文中, 针对现有联邦图学习方法在优化权重感知中学习偏向性不足以及局部信息发散的问题, 提出一种风险感知动态聚合的图联邦学习方法 FedRPDA, 以全面应对数据分布异构性与类别极端不平衡所带来的挑战, 总

体框架如图 2 所示. 该方法围绕两项关键策略展开: 一方面, 典型风险动态聚合策略通过量化客户端本地图中欺诈节点的结构性风险强度来评估其在全局更新过程中的贡献度, 在此基础上结合时间衰减机制的动态权重映射函数调整客户端聚合权重, 使模型在训练初期优先聚焦于具有高风险代表性的客户端 (即数量占多数, 且具备较多风险关联链接的欺诈样本), 从而在数据异构环境下引导全局模型优先学习区分正常样本与典型异常样本的能力. 另一方面, 针对样本稀少但结构多样的非典型欺诈样本, 构建了多样化风险平均聚合策略, 基于本地类别原型提取与全局原型聚合实现跨客户端的类别语义共享, 并结合针对欺诈样本的变分扰动特征增强与全局原型引导的对比学习机制, 提升对多样化欺诈模式的识别能力. 两种策略相辅相成, 共同支撑模型在动态变化与高度不平衡的数据环境下实现鲁棒的欺诈检测性能.

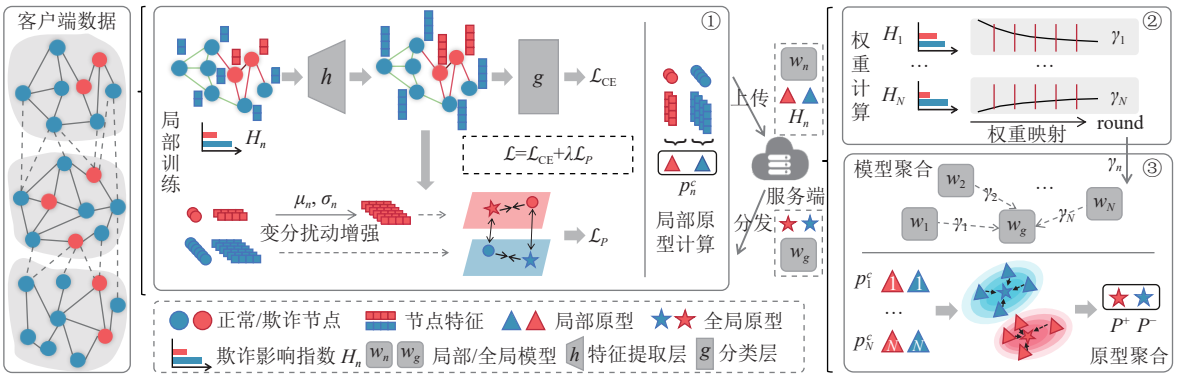


图 2 基于风险感知动态聚合的图联邦学习方法框架

3.1 典型风险动态聚合策略

在经典联邦学习框架中, 客户端的聚合权重通常依据其所持本地数据样本数量确定. 在图结构数据的节点分类任务中, 这一权重往往与客户端图中节点数量成正比. 然而, 在欺诈检测这一特定场景中, 此类加权聚合策略存在显著局限性, 特别是在类别极端不均衡的条件下, 部分客户端虽然拥有较多节点, 但其所含欺诈节点数量极少, 导致训练出的局部模型对欺诈节点的建模能力有限, 却在全局聚合中被赋予较高的权重, 从而削弱了全局模型对欺诈模式的感知能力.

为此, 我们提出了一种基于加权 PageRank 的结构性风险强度评估机制, 用于度量客户端在欺诈检测任务中的实际贡献. PageRank[43]本质上是一种基于随机游走的节点重要性评估方法, 能够反映图中节点的全局结构影响力. 在欺诈检测任务中, 欺诈节点常表现出特定的结构特征, 例如高度互联的欺诈团伙或桥接多个正常节点的欺诈节点. 基于此, 可以在异构拓扑空间中建立欺诈风险传播的量化表示, 更准确地衡量客户端的风险贡献度. 具体而言, 对于每个客户端图 G_n, 我们对其每个节点 $v \in V_n$, 计算 PageRank 值 $PR(v)$:

$$PR(v) = \frac{1-d}{N} + d \sum_{u \in \mathcal{N}(v)} \frac{PR(u)}{|\mathcal{N}(u)|} \tag{2}$$

其中, d 为阻尼因子 (默认设置为 0.85), 控制信息传递的衰减程度; $\mathcal{N}(v)$ 表示节点 v 的邻居集合, $|\mathcal{N}(u)|$ 是邻居 u 的度. 在此基础上, 我们定义客户端 n 的相对欺诈影响指数 H_n 如下:

$$H_n = \frac{\sum_{v \in V_n} PR(v)\mathbb{I}(y_v = 1)}{\sum_{v \in V_n} PR(v)} \tag{3}$$

其中, $\mathbb{I}(y_v = 1)$ 为指示函数, y_v 是节点 v 的标签, $y_v = 1$ 表示该节点为欺诈节点, $y_v = 0$ 表示该节点为正常节点. 据此可以有效规避欺诈节点占比较低所导致的权重失真问题, 同时抑制孤立欺诈节点对全局模型造成的扰动, 更精确地反映客户端在欺诈行为建模中的实际贡献, 在聚合过程中为具有较高结构性风险强度的客户端分配更高的权重.

尽管通过各个客户端的相对欺诈影响指数进行权重分配可以在一定程度上提升模型对欺诈节点的识别能力,

但在联邦学习的训练过程中, 客户端的重要程度并不是始终不变的. 训练初期, 具有较高 H_n 值的客户端通常包含更多、更具代表性的欺诈节点, 其上传的局部模型能够提供丰富的异常模式信息. 此阶段放大这类客户端的聚合权重, 有助于模型快速捕捉到典型欺诈模式, 从而提升整体泛化性能. 随着训练的推进, 局部模型的重要性会逐渐被稀释, 这是因为模型在初始阶段已经通过对高 H_n 值客户端的关注, 获得了较为全面的特征表示, 而后期的训练更多是对已获得的知识进行微调, 在这种情况下, 继续强化个别客户端的贡献反而会干扰模型的稳定性.

为此, 我们设计了一种时间衰减机制的动态权重映射函数, 随通信轮次变化调整客户端聚合权重 γ_n, 具体定义为:

$$\gamma_n = \frac{1}{1+\mathrm{e}^{-\tau(H_n-0.5)}} \tag{4}$$

其中, 映射敏感度参数 τ 随通信轮数 t 增加而逐渐收敛, 其定义如下:

$$\tau = \alpha \mathrm{e}^{-\frac{t}{T}} + \beta \tag{5}$$

其中, t 为当前客户端与服务端通信轮次, T 为总的通信轮次, α 和 β 为超参数, 用于控制初期权重差异的放大程度及后期的平滑收敛趋势. 在训练初期, 较大的 τ 值使得权重函数对 H_n 更为敏感, 强调高结构性风险强度客户端的重要性. 随着通信轮次 t 的增加, τ 逐渐减小, 反映出模型对不同客户端的权重分配趋于平衡. 这样全局模型的聚合过程可以表示为:

$$w_g = \sum_{n=1}^{N} \frac{\gamma_n}{\sum_{n'=1}^{N}\gamma_{n'}} w_n \tag{6}$$

通过上述典型风险动态聚合策略, 全局模型能够在训练初期优先整合具有高欺诈代表性的客户端模型, 有效建立对典型欺诈模式的共性表征, 并随着训练深入逐渐平衡各客户端的影响力, 增强模型在多样环境下的稳定性与泛化能力.

3.2 多样化风险平均聚合策略

由于不同客户端图数据在结构和分布上的高度异构性, 欺诈模式在全局范围内呈现出多样化的特性. 尽管通过典型风险动态聚合策略可使模型优先聚焦于频发且具代表性的典型欺诈模式, 从而构建较强的全局表征能力, 但在面对数量较少且异常特征分布差异显著的非典型欺诈样本时, 模型的感知能力和识别性能仍受到显著制约. 这类异常样本由于数量较少且结构多样, 在数据集中不具备足够的代表性, 导致在训练过程中容易被主流模式所淹没, 难以充分捕捉其特征表达. 为缓解这一问题, 我们提出多样化风险平均聚合策略, 通过引入全局原型引导的对比学习机制, 提升模型对稀有异常样本的判别能力.

考虑到数据隐私约束, 服务端无法直接访问客户端的原始图数据, 因此我们先在客户端本地进行类别原型的计算. 具体地, 对于第 n 个客户端, 其类别 $c \in \{+,-\}$ (分别代表欺诈类别和正常类别) 对应的局部原型向量 p_n^c 定义如下:

$$p_n^c = \frac{1}{\left|V_n^c\right|} \sum_{v \in V_n^c} h(v) \tag{7}$$

其中, V_n^c 表示客户端 n 中类别为 c 的节点集合, $\left|V_n^c\right|$ 为对应的节点数量, $h(\cdot)$ 表示图神经网络的消息传递层, 通过聚合邻居节点的特征信息来更新自身嵌入表示. 局部模型训练结束之后, 各客户端将其计算所得的原型向量与模型参数一同上传至服务端, 服务端依据客户端对应类别的节点数量进行加权聚合, 获得全局类别原型:

$$P^c = \sum_{n=1}^{N} \frac{\left|V_n^c\right|}{\sum_{n'=1}^{N}\left|V_{n'}^c\right|} p_n^c \tag{8}$$

全局原型向量 P^+ 与 P^- 分别为欺诈类别与正常类别的通用语义表示, 后续将其发送给每个客户端用于局部训练过程. 在全局原型知识的指导下, 客户端可通过对比学习机制增强对非典型欺诈样本的建模能力, 进一步优化其在嵌入空间中的可分性. 然而, 我们注意到局部数据的极度不平衡导致欺诈样本在对比学习阶段受到强烈的类间压制, 这不仅使得欺诈节点难以靠近其相应的类原型, 还导致其与正常节点的边界模糊化. 为了平衡类间的排斥强度, 我们设计了基于变分扰动的欺诈类数据增强机制, 利用欺诈样本的局部特征分布生成虚拟样本. 具体而言, 在每轮本地训练的 minibatch 中, 统计该批次内欺诈节点的特征均值 $\mu_{n,b}$ 及方差 $\sigma_{n,b}^2$:

$$\mu_{n,b} = \frac{1}{\left|V_{n,b}^{+}\right|} \sum_{v \in V_{n,b}^{+}} h(v) \tag{9}$$

$$\sigma_{n,b}^{2} = \frac{1}{\left|V_{n,b}^{+}\right|} \sum_{v \in V_{n,b}^{+}} \left(h(v) - \mu_{n,b}\right)^{2} \tag{10}$$

其中, $V_{n,b}^{+}$ 表示客户端 n 在第 b 批次中的欺诈节点集合. 考虑到批次数据的更新导致在损失计算时难以一次性获取所有节点数据, 为了降低内存开销, 我们采用指数加权移动平均方式更新策略平滑统计量:

$$\mu_{n,b} = (1-m)\mu_{n,b-1} + m\mu_{n,b} \tag{11}$$

$$\sigma_{n,b}^{2} = (1-m)\sigma_{n,b-1}^{2} + m\sigma_{n,b}^{2} \tag{12}$$

其中, m 为移动更新的动量超参数, $\mu_{n,b-1}$ 和 $\sigma_{n,b-1}^{2}$ 分别为历史批次的均值和方差, $\mu_{n,b}$ 和 $\sigma_{n,b}^{2}$ 分别为当前批次的均值和方差. 基于以上统计量生成虚拟欺诈节点特征集合:

$$\tilde{p}_{n,b}^{+} = \mu_{n,b} + \sigma_{n,b}\epsilon,\ \epsilon \sim N(0, I) \tag{13}$$

其中, ϵ 为从标准正态分布 $\mathcal{N}(0, I)$ 中采样的噪声, I 为单位矩阵, $\sigma_{n,b}$ 控制扰动幅度. 通过将生产的虚拟欺诈样本与真实欺诈样本节点特征集合并, 形成增强后的表示集:

$$h\left(V_{n,b}^{+}\right) = \left\{\tilde{p}_{n,b}^{+}\right\} \cup \left\{h(v) \mid v \in V_{n,b}^{+}\right\} \tag{14}$$

该增强样本集可与真实欺诈节点共同构成平衡的数据对比集合, 以缓解类别不平衡对训练过程的干扰. 在对比学习阶段, 利用上述增强样本与全局类别原型之间的距离关系构建监督信号, 对于每个节点 v, 其与类别 c 的原型距离定义为:

$$\mathcal{D}_{v}^{c} = \|h(v) - P^{c}\|_{2} \tag{15}$$

其中, $\mathcal{D}_{v}^{+}$ 表示节点与欺诈类别原型的距离, $\mathcal{D}_{v}^{-}$ 表示节点与正常类别原型的距离. 对于欺诈节点应该拉近其与 P^{+} 的距离, 增大其与 P^{-} 的距离, 正常节点则相反. 通过 *Softmax* 函数计算节点与不同类别原型的距离分布, 从而获得归一化概率解释, 并以交叉熵损失进行监督:

$$\mathcal{L}_{P} = -\sum_{v \in V_{n}} \left[y_{v}\log(s_{v}) + (1-y_{v})\log(1-s_{v})\right] \tag{16}$$

$$s_{v} = Softmax\left(-\mathcal{D}_{v}^{(y_{v})}\right) \tag{17}$$

通过增大类间距离和促进类内紧凑, 所提出的多样化风险平均聚合策略显著增强了模型对结构复杂、数量稀少的欺诈节点的表征能力. 这一策略使得这些欺诈样本能够在全局原型知识的指导下在表征空间朝向典型欺诈样本靠近, 从而提升了模型在同类样本上的辨识能力. 此外, 该策略还增强了模型在正常与异常样本之间的区分能力, 确保模型在面对不平衡数据时依然能够保持鲁棒的识别性能.

3.3 训练过程

通过结合典型风险动态聚合策略和多样化风险平均聚合策略, FedRPDA 框架中客户端本地训练和服务端全局聚合流程分别由算法 1 和算法 2 形式化描述.

算法 1. 客户端的局部训练过程.

输入: 局部训练迭代次数 $Iter$, 当前通信轮次 t, 第 t 轮的全局模型参数 w_{g}^{t}, 欺诈节点和正常节点的全局原型 $\{P^{+}, P^{-}\}$, 局部图数据 $G_{n} = (V_{n}, E_{n}, X_{n}, Y_{n})$, 学习率 η;

输出: 局部模型参数 w_{n}^{t}, 局部原型 $\{p_{n}^{+}, p_{n}^{-}\}$, 相对欺诈影响指数 H_{n}.

初始化 $w_{n}^{t} \leftarrow w_{g}^{t}$;

1. for $epoch = 1, 2, \ldots, Iter$ do

2. $\quad \mathcal{L}_{\mathrm{CE}} \leftarrow CrossEntropyLoss(f(w_{n}, G_{n}), Y_{n})$; // 计算节点交叉熵分类损失

3. if $t > 1$ then
4. 根据公式 (13)、公式 (14) 对欺诈节点通过变分扰动进行数据增强;
5. 根据公式 (15) 计算节点嵌入与$\{P^+, P^-\}$的距离;
6. 根据公式 (16)、公式 (17) 计算原型对比损失 $\mathcal{L}_P$;
7. else
8. $\mathcal{L}_P = 0$;
9. 根据公式 (2)、公式 (3) 计算相对欺诈影响指数 H_n;
10. end if
11. $\mathcal{L}_n = \mathcal{L}_{\mathrm{CE}} + \lambda\mathcal{L}_P$; // 计算总损失
12. $w_n^t \leftarrow w_n^t - \eta\nabla\mathcal{L}_n$; // 更新模型参数
13. end for
14. 根据公式 (7) 分别计算局部原型$\{p_n^+, p_n^-\}$;
15. Return w_n^t, p_n^+, p_n^-, H_n;

算法 2. 服务端的全局聚合过程.

输入: 当前通信轮次 t, 客户端模型参数集合 $\{w_1^t, w_2^t, \ldots, w_N^t\}$, 客户端相对欺诈影响指数集合 $\{H_1, H_2, \ldots, H_N\}$, 客户端类别原型集合 $\{p_1^c, p_2^c, \ldots, p_N^c\}$;
输出: 全局模型参数 w_g^{t+1}, 全局原型$\{P^+, P^-\}$.

1. 接收来自所有客户端的本地模型参数 $\{w_n^t\}_{n=1}^N$、局部类别原型 $\{p_n^c\}_{n=1}^N$ 以及相对欺诈影响指数 $\{H_n\}_{n=1}^N$;
2. 利用 $\{H_1, H_2, \ldots, H_N\}$ 根据公式 (6) 计算客户端的聚合权重 γ_n;
3. 根据权重 γ_n 更新全局模型参数 $w_g^{t+1} \leftarrow \sum_{n=1}^{N} \frac{\gamma_n}{\sum_{n'=1}^{N}\gamma_{n'}} w_n^t$;
4. 根据公式 (11)、公式 (12) 计算全局原型$\{P^+, P^-\}$;
5. Return w_g^{t+1}, P^+, P^-;

在初始化阶段, 每个客户端接收服务端下发的全局模型初始参数 w_g^0, 由于尚未建立跨客户端共享的类别语义信息, 当前阶段不涉及原型对比相关操作. 因此在首轮本地训练过程中, 客户端仅基于本地的图结构数据 G_n 执行节点分类任务, 利用交叉熵损失 $\mathcal{L}_{\mathrm{CE}}$ 对局部模型参数进行更新. 在此阶段, 客户端评估图数据的结构性风险强度并结合节点的类别标签计算 H_n, 用于刻画客户端在全局欺诈建模中的潜在贡献度. 在随后的通信轮次中, 各个客户端除了会接收到来自服务端返回的全局模型参数 w_g^t, 还会接收到全局正常类原型 P^- 和欺诈类原型 P^+. 在这一训练过程中, 将采用多样化风险平均聚合策略对 minibatch 的欺诈节点特征进行变分扰动增强以缓解类别不平衡对模型训练的不利影响, 并利用全局原型知识构建对比损失 $\mathcal{L}_P$ 指导结构多样且分布稀疏的非典型欺诈节点在嵌入空间中聚合至共性异常表示, 提升模型对这些样本的学习能力. 此时的训练目标函数由用于节点分类任务的交叉熵损失和原型对比损失共同计算:

$$L = \mathcal{L}_{\mathrm{CE}} + \lambda\mathcal{L}_P \tag{18}$$

其中, λ 为控制交叉熵损失与原型对比损失之间权重关系的平衡因子. 在每一轮本地训练结束后, 客户端会利用本地节点数据计算局部类别原型 p_n^c, 并将其与更新后的局部模型参数 w_n^t 以及相对欺诈影响指数 H_n 一并上传到服务端.

在服务端聚合阶段, 服务器首先接收来自各客户端的上传信息. 根据客户端的相对欺诈影响指数集合 $\{H_n\}_{n=1}^N$, 服务端通过带有时间衰减机制的动态权重映射函数计算每个客户端的聚合权重 γ_n, 以实现在训练早期强化对高风险客户端的关注、训练后期逐步趋于权重平衡的动态聚合策略. 在此基础上, 服务端对客户端上传的模型参数执

行加权聚合, 更新得到新一轮全局模型参数 w_g^{t+1}. 随后, 依据客户端上传的局部类别原型加权聚合更新全局类别原型表示 $\{P^+, P^-\}$. 完成更新后, 服务端将模型参数和类别原型广播至各客户端, 用于启动下一轮本地训练.

3.4 收敛性分析

本节对于 FedRPDA 在非凸条件下的收敛性给出了简要的理论分析, 并作出与现有框架相似的假设条件.

假设 1. L-光滑性. 每个客户端的局部目标函数 $\mathcal{L}_n$ (公式 (18)) 是 L-光滑的, 即对任意模型参数 w_n, w_n', 满足:

$$\|\nabla\mathcal{L}_n(w_n') - \nabla\mathcal{L}_n(w_n)\| \leqslant L\|w_n' - w_n\| \tag{19}$$

等价于二次上界条件:

$$\mathcal{L}_n(w_n') \leqslant \mathcal{L}_n(w_n) + \langle\nabla\mathcal{L}_n(w_n), w_n' - w_n\rangle + \frac{L}{2}\|w_n' - w_n\|^2 \tag{20}$$

假设 2. 无偏梯度. 基于小批量数据 ξ 计算的随机梯度 $\nabla\mathcal{L}_n(w_n;\xi)$ 是对真实梯度的无偏估计:

$$\mathbb{E}_\xi[\nabla\mathcal{L}_n(w_n;\xi)] = \nabla\mathcal{L}_n(w_n) \tag{21}$$

假设 3. 有界方差. 随机梯度方差有界, 即存在常数 σ_n^2, 使得对任意 w_n 有:

$$\mathbb{E}_\xi\left[\|\nabla\mathcal{L}_n(w_n;\xi) - \nabla\mathcal{L}_n(w_n)\|^2\right] \leqslant \sigma_n^2 \tag{22}$$

假设 4. 有界非相似性. 存在常数 $\beta^2 \geqslant 1$ 与 $\kappa^2 \geqslant 0$, 使得对任意模型参数 w_n 有:

$$\sum_{n=1}^{N}\rho_n\|\nabla\mathcal{L}_n(w_n)\|^2 \leqslant \beta^2\|\sum_{n=1}^{N}\rho_n\nabla\mathcal{L}_n(w_n)\|^2 + \kappa^2 \tag{23}$$

其中, $\rho_n = \dfrac{\gamma_n}{\sum_{n'=1}^{N}\gamma_{n'}}$ 表示归一化聚合权重.

假设 5. Lipschitz 连续. 节点特征提取网络 $h(\cdot)$ 满足 L_2-Lipschitz 连续:

$$\|h(w_n';v) - h(w_n;v)\| \leqslant L_2\|w_n' - w_n\|,\ \forall v \in V_n \tag{24}$$

定理 1. FedRPDA 的非凸收敛性. 基于上述假设, 在非凸条件假设下, FedRPDA 算法经过 T 轮通信后可以达到如下收敛保证:

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\|\nabla\mathcal{L}\left(w_g^t\right)\|^2\right] \leqslant \frac{4\beta^2}{\eta KT}\left(\mathcal{L}\left(w_g^0\right) - \mathcal{L}^*\right) + 2\beta^2L^2\eta^2K\left(D^2+\sigma^2\right) + \frac{2\kappa^2}{K} + 2\beta^2L\eta\sigma^2 + 4\beta^2\lambda L_2G \tag{25}$$

证明: 考虑在通信轮次 t 时客户端 n 的第 k 步本地更新:

$$w_{n,k+1}^t = w_{n,k}^t - \eta\nabla\mathcal{L}_n\left(w_{n,k}^t;\xi_k\right) \tag{26}$$

根据假设 1–3 及假设 5, 可以得到任意客户端在单个通信轮次的局部目标函数偏差界限:

$$\mathbb{E}\left[\mathcal{L}_n\left(w_{n,K}^t\right)\right] \leqslant \mathcal{L}_n\left(w_g^t\right) - \sum_{k=0}^{K-1}\left(\eta - \frac{L\eta^2}{2} - \frac{\lambda}{2}\right)\left\|\nabla\mathcal{L}_n\left(w_{n,k}^t\right)\right\|^2 + \frac{LK\eta^2}{2}\sigma_n^2 + \lambda\eta L_2KG \tag{27}$$

通过选取适当的 η 和 λ 值, 且扰动项受控, 可以保证单轮期望下降.

对于服务端加权聚合项, 可以得到全局损失的递推上界:

$$\mathbb{E}\left[\mathcal{L}\left(w_g^{t+1}\right)\right] \leqslant \sum_{n=1}^{N}\rho_n^t\left(\mathcal{L}_n\left(w_g^t\right) - \frac{\eta}{2}\sum_{k=0}^{K-1}\left\|\nabla\mathcal{L}_n\left(w_{n,k}^t\right)\right\|^2 + \frac{LK\eta^2}{2}\sigma_n^2 + \lambda\eta L_2KG\right) \tag{28}$$

由假设 4 和梯度有界性 $D \triangleq \max_{w,n}\|\nabla\mathcal{L}_n(w_n)\|$, 对 $t = 0$ 到 $T-1$ 求和, 可以得到:

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\left\|\nabla\mathcal{L}\left(w_g^t\right)\right\|^2\right] \leqslant \frac{4\beta^2}{\eta KT}\left(\mathcal{L}\left(w_g^0\right) - \mathcal{L}^*\right) + 2\beta^2L^2\eta^2K\left(D^2+\sigma^2\right) + \frac{2\kappa^2}{K} + 2\beta^2L\eta\sigma^2 + 4\beta^2\lambda L_2G \tag{29}$$

其中, $\mathcal{L}^* \triangleq \min_w\mathcal{L}(w)$.

这表明在非凸情形下, 随着通信轮次 T 的增加, FedRPDA 的平均梯度范数平方持续下降, 其收敛水平受异构性、噪声以及多样化项共同约束. 详细推导过程可见附录 A.

4 实验分析

为验证所提出方法的有效性, 本节在多个真实欺诈场景下开展实验评估. 首先, 选取 3 个具代表性的欺诈检测数据集构成全局图, 并采用广泛使用的子图划分策略为各客户端分配局部图数据. 随后, 介绍对比的基线方法和具体的实验设置. 在此之后, 我们将通过实验回答以下 3 个问题.

(1) 问题 1. 与现有先进的联邦学习方法相比, FedRPDA 是否能实现更优的检测性能?

(2) 问题 2. 在方法有效的前提下, 其性能提升从何而来?

(3) 问题 3. FedRPDA 相较于其他先进方法有什么明显优势?

4.1 实验数据

我们在 3 个典型的欺诈检测场景下选取了广泛使用的公开数据集, 表 1 提供了关于数据集的详细信息.

表 1 数据集详细信息

数据集	欺诈场景	节点数	边数	节点特征数	欺诈节点数	欺诈率 (%)
T-Finance	金融交易网络	39 357	21 222 543	10	1 804	4.58
T-Social	社交网络	5 781 065	73 105 508	10	174 280	3.01
Amazon	评论系统	11 944	4 398 392	25	821	6.87

T-Finance 数据集[44]主要用于金融交易网络中的异常用户检测. 该数据集包含了大量真实交易记录, 节点表示匿名用户, 具有包括注册天数、交易行为和日志活动在内的 10 个维度的特征. 如果两个用户之间存在交易关系, 在图中添加边连接. 对于存在欺诈、洗钱和在线赌博行为的用户, 将其标记为异常.

T-Social 数据集[44]来自真实社交媒体平台, 用于检测虚假用户或异常行为, 与 T-Finance 数据集类似, 图中的每个节点表示一个匿名用户, 当两个用户的好友关系持续超过 3 个月时在图中连接边. 该数据集是一个超大规模的图, 具有 578 万个节点和 7 000 多万条边, 对于欺诈检测任务具有很大的挑战性.

Amazon 数据集[45]用于识别评论系统中的异常用户. 该数据集由亚马逊公司所提供, 包含乐器类别下的产品评论信息. 其中每个节点表示一个用户, 若用户评论被判定为有用的比例低于 20%, 则标记为欺诈实体, 若高于 80%, 则视为正常用户. 此外, 该数据集还包含多种用户之间的关系类型, 包括评论同一商品的 U-P-U 边、在一周内给出相同评分的 U-S-U 边以及评论相似度排名前 5% 的 U-V-U 边, 构成了一个具有多关系结构特征的异构用户交互图.

基于上述 3 个数据集, 为模拟真实场景中不同机构间的数据隔离与结构异构性, 我们采用在图联邦学习领域中广泛使用[17,26,46]的 Louvain 社区划分算法[47], 对每个数据集的全局图进行划分, 其中分辨率参数设置为 1.0 以控制划分的粒度, 划分后删除跨子图连接边, 确保各客户端子图的结构独立性. 节点特征在建模前采用 Z-score 标准化处理, 未进行边裁剪或特征降维操作, 以保持图结构与特征信息的完整性. 需要说明的是, 尽管各子图在规模与结构上存在差异, 但其在欺诈风险行为的判定性上是一致的. 这是因为各数据集的欺诈标签均来源于统一的业务规则与风险识别标准, 确保了正常节点与欺诈节点的定义在全局范围内保持一致.

4.2 基准方法

目前, 针对欺诈检测任务的图联邦学习研究仍处于起步阶段, 尚缺乏统一的评估标准和广泛认可的基线方法. 因此, 本文选取了在其他联邦学习场景以及通用图节点分类任务中表现优异的几种具有代表性的方法进行对比分析, 以全面评估所提出方法的有效性与鲁棒性.

以经典的 FedAvg[11]作为基础方法, FedProx[14]、MOON[15]和 FedProc[48] 分别通过引入不同的训练约束机制来缓解由数据异构性引发的模型偏移问题, 并在计算机视觉等通用联邦学习任务中取得了良好表现. 此外, 基于 Shapley[21]值的客户端贡献评估方法提供了一种公平且具解释性的聚合权重分配策略, 能够精确量化每个客户端对全局模型性能的实际贡献, 从而优化联邦聚合效果. 在联邦图学习中, FedSage+[17]和 FedSSL[37] 则通过结合图结构信息改进训练与聚合过程, 提升了模型在图数据处理中的表达能力与鲁棒性.

4.3 评价指标

考虑到图欺诈检测任务中普遍存在的严重类别不平衡问题, 仅依赖准确率作为性能评估指标往往难以全面反映模型在欺诈检测方面的性能 (模型将所有节点预测为正常节点时准确率依然较高). 因此, 本文采用以下 4 项指标对模型的检测效果进行全面评估.

Recall, 即召回率, 衡量模型识别实际欺诈样本的能力, 计算公式表示为:

$$Recall = \frac{TP}{TP+FN} \tag{30}$$

其中, TP (true positive) 表示真正例, FN (false negative) 表示假负例, *Recall* 表示模型能从所有真实欺诈样本中正确识别出的比例.

$F1$-macro 对各类别分别计算 $F1$ 分数并取平均, 适用于类别不平衡的场景. $F1$ 值综合考虑了精确率 (*Precision*) 与召回率, 定义如下:

$$Precision = \frac{TP}{TP+FP} \tag{31}$$

$$F1 = 2\times\frac{Precision\times Recall}{Precision+Recall} \tag{32}$$

其中, FP (false positive) 表示假正例. $F1$-macro 在计算时对每个类别赋予相同的权重, 不受样本数量的影响, 适合于类别不平衡的任务中.

AUC (area under curve) 被定义为 ROC (receiver operating characteristic curve) 曲线下与坐标轴围成的面积, ROC 曲线通过改变分类阈值, 表示真阳性率 (true positive rate, TPR) 和假阳性率 (false positive rate, FPR) 的关系, 衡量模型在不同阈值下对正负样本的区分能力.

AP (average precision) 通过计算 Precision-Recall (P-R) 曲线下的面积来衡量模型性能, P-R 曲线通过改变分类阈值, 综合评估模型在不同阈值下的精度与召回性能, 有效地衡量误报率, 更准确地反映模型对欺诈节点的识别能力.

上述所有评价指标的范围都在 0–100 (以百分数的形式呈现), 值越高表示模型效果越好.

4.4 实验设置

本文实验在所有数据集上均使用 GraphSAGE[49] 模型作为基础图神经网络框架, 该模型通过有效捕捉节点的局部结构特征表示, 能够适应不同规模和结构的图数据, 在图联邦学习中被广泛使用[17,40,42]. 在联邦学习的设置中, 客户端与服务端的最大通信轮次设置为 200, 客户端数量为 4, 本地迭代次数为 5. 优化器使用 Adam, 学习率为 1E–3, 权重衰减设置为 5E–4. 在本地模型训练阶段, 采用 minibatch 方式进行模型更新, T-Finance 和 Amazon 数据集的批量大小设置为 1 024, 由于 T-Social 数据集数据规模较大, 批量大小设置为 10 240, 以保证训练的收敛效率与稳定性. 在超参数设置方面, 采用网格搜索策略确定关键参数的最优取值, 对损失平衡因子 λ 在 $\{0.1, 0.2, 0.5, 1, 2, 5\}$ 中进行搜索, 权重映射函数中的参数 α 和 β 分别在 $\{5, 6, 7, 8\}$ 和 $\{2, 3, 4, 5\}$ 中搜索, 移动更新的动量超参数 m 默认设置为 0.1. 此外, 为保证对比结果的一致性和公正性, 所有实验均在同一硬件平台上运行, 实验环境配置为 NVIDIA RTX A6000 GPU 和 Intel(R) Xeon(R) Platinum 8260M CPU @ 2.30 GHz, 对所有模型训练过程设置固定随机种子.

4.5 实验结果与分析

为回答问题 1, 本节对 FedRPDA 的检测性能进行全面评估, 并与多种先进联邦学习方法进行对比, 实验结果如表 2 所示. 接下来, 对实验结果进行详细分析.

(1) 总体来看, FedRPDA 在所有数据集和评估指标上均优于现有基线方法, 展现出优异的检测能力与泛化性能. 具体而言, FedRPDA 相较于 FedAvg, 在 T-Finance、T-Social 和 Amazon 数据集上的 $F1$-macro 分别提升了 20.26%、6.72% 和 0.42%, *Recall* 分别提升了 20.62%、26.09% 和 2.56%, AUC 分别提高了 1.21%、0.95% 和 0.34%, AP 分别提高了 3.33%、9.86% 和 2.41%. 可以发现, 虽然一些现有的方法在 Amazon 数据集上可以取得较为不错的效果, 这主要是因为该数据集中欺诈节点所占比例较大, 且数据规模相对较小, 导致在不同客户端的数据

异构性较低, 使得模型能够有效学习到欺诈节点的特征. 然而, 在欺诈节点比例最小且图结构最复杂的 T-Social 数据集上, 除 Shapley 方法外大多数现有方法性能提升有限, 这表明现有方法在面对高度不平衡的复杂图结构数据时适应性存在局限. 而 Shapley 方法在模型性能上所带来的提升是以牺牲时间效率为代价换来的, 其高达 O(N!) 的算法复杂度和对服务端设置辅助测试集的需求, 在一些实际的联邦学习场景中难以应用. 相比之下, 本文所提出的方法在各个场景下的数据集都表现出优异的性能, 尤其是在对欺诈节点的正确识别中, 展现出了强大的能力.

表 2 FedRPDA 与基准方法性能比较 (%)

方法	T-Finance				T-Social				Amazon			
	*F*1-macro	*Recall*	AUC	AP	*F*1-macro	*Recall*	AUC	AP	*F*1-macro	*Recall*	AUC	AP
FedAvg	66.53	60.63	94.37	78.27	61.69	72.05	94.16	64.19	92.05	89.70	97.20	87.67
FedProx	72.45	65.28	94.56	79.32	62.13	73.85	94.27	66.59	91.30	90.67	96.85	88.69
MOON	74.94	67.43	94.24	79.25	62.39	76.06	94.36	67.54	91.90	90.81	<u>97.46</u>	89.25
FedProc	70.57	64.21	94.17	78.16	61.03	72.16	93.02	61.89	91.18	90.38	96.16	89.16
Shapley	<u>78.72</u>	<u>71.04</u>	<u>94.61</u>	<u>79.38</u>	<u>64.21</u>	<u>84.38</u>	<u>94.81</u>	<u>71.32</u>	<u>92.29</u>	<u>91.87</u>	97.32	<u>89.72</u>
FedSage+	68.66	62.22	94.10	78.29	62.54	72.71	94.31	64.72	92.13	90.45	97.14	88.22
FGSSL	57.66	54.79	93.81	76.71	49.23	53.29	91.14	43.80	90.82	88.43	96.03	87.17
FedRPDA	**86.79**	**81.25**	**95.58**	**81.60**	**68.41**	**88.14**	**95.11**	**74.05**	**92.47**	**92.26**	**97.54**	**90.08**
提升	+8.07	+10.21	+0.97	+2.22	+4.20	+3.76	+0.30	+2.73	+0.18	+0.39	+0.08	+0.36

注: 加粗表示最优结果,下划线表示次优结果

(2) FedRPDA 在 *F*1-macro 和 *Recall* 指标上得到了显著提升, 这主要得益于风险感知动态聚合策略的特性. 该策略通过聚焦结构性风险较高的客户端, 促使模型在训练初期优先学习典型欺诈样本的共性特征, 并将其转化为全局知识, 引导模型更有效地识别稀有欺诈样本. 同时, 对比学习机制通过缩小类内方差、扩大类间距离, 有效增强了模型对欺诈类别的判别能力, 在提升召回率的同时控制了误报风险, 从而提高了 *F*1 分数. 此外, 在不同方法的比较中, 我们发现 AUC 指标普遍偏高, 这主要是因为 AUC 不受样本不平衡比例的影响, 反映的是模型区分正负样本的能力, 更关注模型在不同阈值下的整体分类性能. 大多数模型在训练过程中能够较为有效地捕捉到样本的分布特征, 因此即使在某些情况下, 模型的精确率和召回率表现不理想, 但由于整体区分能力较强, AUC 依然表现良好. 相比之下, AP 值对模型在识别稀少类别时的表现非常敏感, 若模型无法有效识别欺诈样本, 或者在提高对欺诈样本识别能力的过程中导致了假阳性 (将正常样本误判为欺诈样本) 的增加, AP 值都会受到显著影响, 这使得其更能反映模型在实际欺诈检测场景中的能力. 在本文的方法中, AP 指标的提升幅度普遍超过 AUC, 表明 FedRPDA 在增强对稀有欺诈样本检测能力的同时, 仍能维持较高的正常样本识别准确率, 体现出良好的实际应用价值.

(3) 进一步评估 FedRPDA 在非图结构欺诈检测方法上的泛化能力, 我们在保持联邦框架与训练机制不变的前提下, 将原有的 GraphSAGE 架构替换为多层感知机 (MLP), 并仅利用图中节点的属性特征作为输入, 相关实验结果如表 3 所示. 需要说明的是, 由于 T-Social 数据集高度依赖图结构, 即使在集中式环境下采用非图方法也难以获得有效训练, 因此在本实验中不纳入该数据集的评估. 从结果来看, 非图结构的欺诈检测方法尽管会有更低的计算复杂度, 但整体检测性能明显低于完整图结构下的图神经网络方法. 这表明在欺诈节点稀缺、模式复杂的环境中, 图结构所提供的上下文信息对于建模节点间的潜在关系与判别风险特征具有关键作用. 值得注意的是, 即使在缺乏结构信息的情境下, FedRPDA 仍然取得了优于其他方法的性能表现, 验证了其两项核心策略在非结构输入下依然具备良好的泛化能力.

表 3 在非图结构欺诈检测方法下 FedRPDA 与基准方法性能比较 (%)

方法	T-Finance				Amazon			
	*F*1-macro	*Recall*	AUC	AP	*F*1-macro	*Recall*	AUC	AP
FedAvg	51.41	51.30	92.41	67.71	91.46	87.20	96.90	87.16
MOON	52.71	52.00	92.54	67.95	91.63	90.05	97.00	87.21
Shapley	53.35	52.34	92.75	68.14	91.70	89.64	97.09	88.01
FedRPDA	**55.69**	**53.66**	**92.86**	**68.62**	**91.77**	**90.42**	**97.39**	**88.30**

注: 加粗表示最佳结果

4.6 消融分析

为了回答问题 2, 本节以 T-Finance 数据集为例, 对 FedRPDA 中各项关键策略对模型性能的贡献进行消融分析. 具体实验结果如表 4 所示, 具体分析如下.

表 4 T-Finance 数据集上的消融实验结果 (%)

索引	典型风险动态聚合		多样化风险平均聚合		*F*1-macro	*Recall*	AUC	AP
	w/o 权重映射	w/ 权重映射	w/o 变分扰动增强	w/ 变分扰动增强				
1	—	√	—	√	86.79	81.25	95.58	81.60
2	—	√	—	—	84.59	77.50	94.88	81.01
3	—	—	—	√	83.99	76.74	94.87	79.92
4	√	—	—	—	69.20	62.50	94.52	78.95
5	—	—	√	—	65.66	60.00	94.23	77.86
6	—	—	—	—	66.53	60.63	94.37	78.27

注: 索引6表示不应用两种策略的基准FedAvg方法

(1) 对于典型风险动态聚合策略, 模型性能的提升主要来源于对客户端中欺诈节点结构性风险强度的准确度量, 以及在训练过程中对聚合权重的动态调节机制. 观察表 4 中的实验结果, 与基线方法 FedAvg (索引 6) 相比, 基于客户端重要性进行全局模型加权聚合 (索引 4) 能够一定程度上提升模型效果, 而在此基础上结合具有时间衰减特性的动态权重映射机制 (索引 2) 后模型性能得到显著提升. 这说明, 基于欺诈风险的重要性评估比单纯依赖节点数量的加权策略更能凸显关键客户端的价值. 而根据训练阶段动态调整各客户端的影响力, 使模型在早期更关注具有代表性的高风险客户端, 可以加速对典型欺诈行为的学习, 提升整体识别准确性与训练效率.

(2) 对于多样化风险平均聚合策略, 模型性能的提升主要来源于结合变分扰动的欺诈节点特征增强机制与全局原型引导的对比学习策略的优化. 通过对表 4 索引 3、5、6 的对比可以观察到, 如果在对比学习过程中直接对所有节点施加约束反而会导致负作用, 这是因为严重的类别不平衡导致欺诈节点在对比学习过程中受到类间压制, 影响了模型对于欺诈节点的有效学习. 而当结合基于欺诈节点特征分布生成的虚拟欺诈节点来平衡类别样本数量后, 全局模型性能明显上升. 在全局原型知识的指导下, 一方面各个客户端建立平衡的正常节点和欺诈节点集合, 在特征空间靠近各自类别的原型并远离相反类型的原型可以减轻各个客户端在特征分布上的异构性. 另一方面, 如图 3 所示, 在数据中的结构多样但数量稀少的欺诈样本可以在全局欺诈原型知识的引导下, 向共性欺诈特征靠近, 且样本的特征空间变得更加可区分.

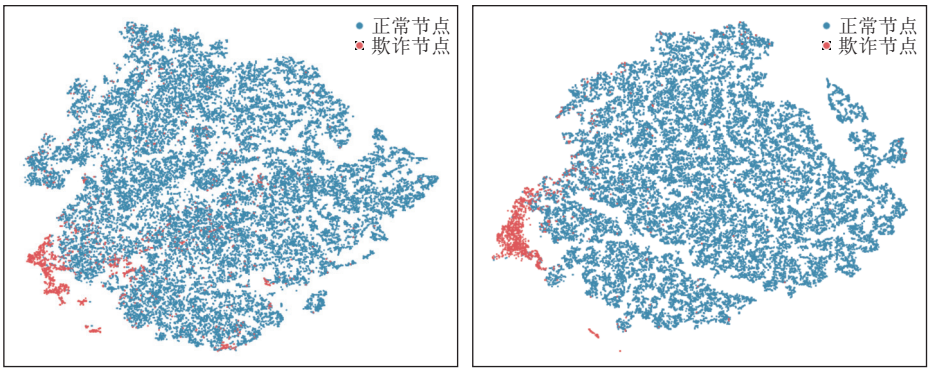


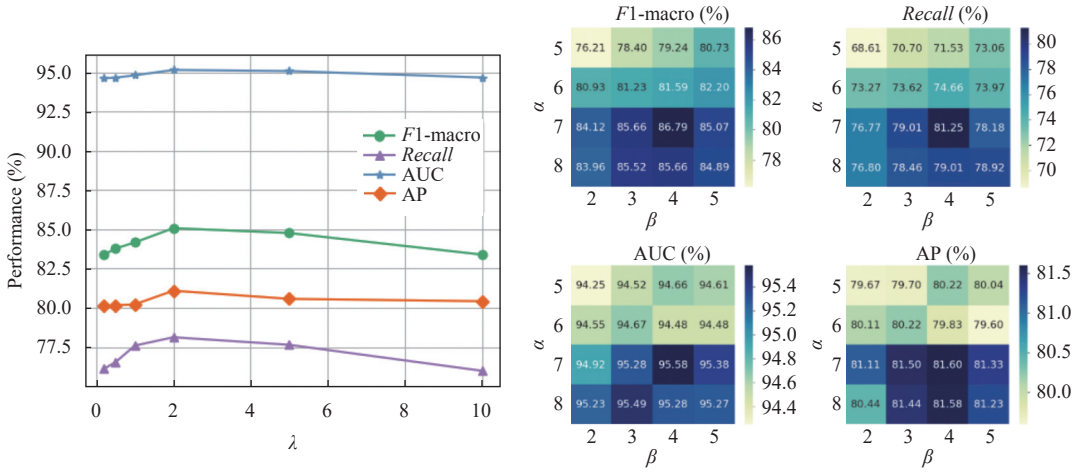
图 3 FedAvg 与 FedRPDA 训练节点特征空间对比

(3) 从索引 1 (同时采用两项策略) 与其他单独策略组合的对比结果可见, 典型风险动态聚合策略与多样化风险平均聚合策略在目标与机制上互为补充. 前者通过客户端聚合权重量化和动态调整帮助模型在复杂环境中快速适应和提升对典型欺诈样本的识别能力, 后者则通过跨客户端的类别语义共享与欺诈节点特征增强的数据平衡机制提升模型对异常风险的整体感知能力, 使其在面临多样化风险时更为鲁棒. 通过结合两种方法, 即本文所提出的

风险感知动态聚合方法, 可以在保证训练稳定性的同时, 有效提升模型在复杂、不平衡数据环境下的欺诈检测能力, 在仅增加线性计算开销的情况下带来显著性能增益.

4.7 超参数影响分析

对于本文方法在不同数据集上的性能评估, 我们通过网格搜索方式对关键超参数进行调优. 图 4(a) 展示了在 T-Finance 数据集上, 平衡因子 λ 对模型性能的影响. 实验结果表明, 当 λ 取值为 2 时, 模型性能达到最佳拐点, 说明该值在平衡原型对比损失与交叉熵损失之间具有良好的协调作用. 图 4(b) 进一步分析了时间动态加权机制中的超参数 α 与 β 对模型效果的影响, 结果显示, 当 α 设为 7、β 设为 4 时, 模型在各项评估指标上均表现最优. 为了保证模型在不同欺诈检测场景下的稳健性, 我们在 T-Social 与 Amazon 数据集上也采用了相同的参数搜索策略, 从而确保在不同数据分布条件下取得最优性能.



(a) 参数 λ 对模型性能的影响 (b) 参数 α、β 对模型性能的影响

图 4 T-Finance 数据集上超参数对模型性能的影响

4.8 训练效率分析

为了回答问题 3, 本节针对 T-Finance 数据集和基准方法在收敛性能、训练时间和系统资源开销上进行了讨论. 在收敛性能方面, 我们记录了各方法在通信轮次与模型 AP 值之间的变化关系, 如图 5 所示. 结果显示, 本文方法与 MOON 和 Shapley 在训练初期均展现出较快的性能提升趋势, 其中 FedRPDA 在多样化风险平均聚合策略的作用下, 不仅在早期实现了快速增益, 还在训练后期维持了较高的检测性能, 并最终实现稳定收敛. 为进一步比较训练效率, 我们统计了各方法达到与 FedAvg 在 100 轮训练中所取得性能相当的 AP 值所需的通信轮次, 结果见表 5.

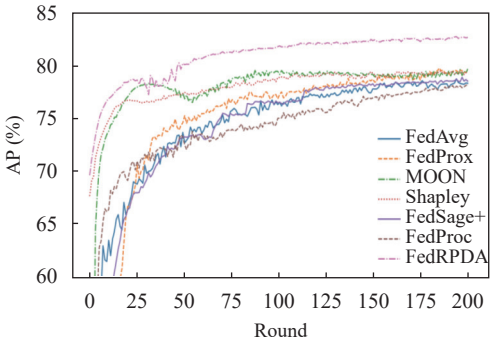


图 5 不同通信轮数下各个方法的 AP 值比较

表 5 各方法收敛效率分析

方法	轮次	加速
FedAvg	100	1.0×
FedProx	73	1.4×
MOON	19	5.3×
FedProc	136	0.7×
Shapley	18	5.6×
FedSage+	84	1.2×
FGSSL	—	—
FedRPDA	**11**	**9.1×**

可以观察到, FedRPDA 仅需 11 轮通信即可达到相同性能, 相较于 FedAvg 显著加速了训练过程, 实现了约 9.1 倍的效率提升. 此外, 为量化模型的系统开销, 我们进一步对比了各方法在单轮训练过程中的资源消耗, 包括客户端与服务端的内存占用以及计算时间复杂度, 相关结果如表 6 所示. 可以观察到, FedRPDA 在 FedAvg 的基础上仅引入了线性级别的额外开销, 其计算成本远低于如 Shapley 等复杂度较高的策略. 上述结果表明, FedRPDA 在有效提升检测性能的同时, 保持了较低的系统资源消耗与良好的扩展性, 具备较强的实际应用潜力.

表 6 各方法在单轮训练所引入的内存和时间开销对比

方法	客户端内存复杂度	服务端内存复杂度	客户端计算复杂度	服务端计算复杂度
FedAvg	$O((b+k)f+f^2)$	$O(Nf^2)$	$O(kmf+nf^2)$	$O(N)$
FedProx	$O((b+k)f+\omega f^2)$	$O(Nf^2)$	$O(kmf+nf^2+f^2)$	$O(N)$
MOON	$O((b+k)f+Qf^2)$	$O(Nf^2)$	$O(kmf+nf^2+Qnf)$	$O(N)$
FedProc	$O((b+k)f+f^2+cf)$	$O(Nf^2+Ncf)$	$O(kmf+nf^2+cf^2)$	$O(N)$
Shapley	$O((b+k)f+f^2)$	$O(N!f^2)$	$O(kmf+nf^2)$	$O(N!)$
FedSage+	$O(L(n+sg)f+f^2)$	$O(LtNf^2)$	$O(L((m+sg)f+(n+sg)f^2))$	$O(N)$
FGSSL	$O(Q(b+k)f+f^2)$	$O(Nf^2)$	$O(Qkmf+Qnf^2)$	$O(N)$
FedRPDA	$O((b+k)f+f^2+cf)$	$O(Nf^2+Ncf)$	$O(kmf+nf^2+cf^2+bf)$	$O(N)$

注: b 表示mini-batch大小, k 表示特征传播步数, n 表示节点数, m 表示边数, c 表示类别数, f 表示特征维度, N 表示客户端数量, ω 表示模型对齐损失项, Q 表示对比学习的查询集大小, s 表示增强的节点数, g 表示生成的邻居数, L 表示模型层数

5 总 结

本文重点关注图联邦学习中的欺诈检测问题, 探讨如何在保障数据隐私的前提下, 通过跨机构合作应对图数据分布高度不平衡和机构间数据异构性带来的挑战, 特别是对多样化风险异常样本识别问题. 为此, 本文提出了一种面向欺诈检测的风险感知动态聚合图联邦学习方法 FedRPDA. 该方法通过引入典型风险动态聚合策略和多样化风险平均聚合策略, 实现对不同欺诈风险形态的协同建模, 前者在数据异构性的环境下增强模型对典型欺诈样本的识别能力, 后者引导非典型欺诈样本在表征空间中向共性异常靠拢, 实现对复杂环境中多样化欺诈行为的有效建模与识别. 在 3 个具有代表性的欺诈检测数据集上的实验结果表明, FedRPDA 相较于现有主流图联邦学习方法在多项评估指标上均实现了性能提升, 特别是在数据分布极度不均衡和大规模场景下, 展现出更优越的异常识别能力. 尽管 FedRPDA 在多个真实欺诈检测场景中表现出良好的鲁棒性和泛化能力, 但方法本身仍存在一定的适用边界. 当前框架假设所有客户端均为诚实参与者, 未考虑潜在的投毒攻击等恶意行为, 在开放或不完全可信的协作环境中可能对系统稳定性构成威胁; 此外, 当本地图数据存在严重标签缺失时, 基于结构性风险评估的动态聚合策略会影响聚合效果. 未来工作可进一步结合鲁棒聚合机制与异常客户端检测策略, 提升方法在复杂协同环境下的安全性, 同时进一步拓展至更广泛的图学习任务, 以增强方法的通用性与应用范围.

References

[1] Song LY, Ma ZY, Li ZH, Shang XQ. Review on temporal graph neural networks for financial risk prediction. Ruan Jian Xue Bao/Journal of Software, 2024, 35(8): 3897–3922 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/7087.htm [doi: 10.13328/j.cnki.jos.007087]

[2] Kurshan E, Shen HD. Graph computing for financial crime and fraud detection: Trends, challenges and outlook. Int'l Journal of Semantic Computing, 2020, 14(4): 565–589. [doi: 10.1142/S1793351X20300022]

[3] Li PC, Li CT. TCGNN: Text-clustering graph neural networks for fake news detection on social media. In: Proc. of the 28th Pacific-Asia Conf. on Knowledge Discovery and Data Mining. Taipei: Springer, 2024. 134–146. [doi: 10.1007/978-981-97-2266-2_11]

[4] Kumar S, Shah N. False information on web and social media: A survey. arXiv:1804.08559, 2018.

[5] Yin M, Qiao S, Chen W, Jiang JJ. Collective emotional stabilization method for social network rumor detection. Ruan Jian Xue

Bao/Journal of Software, 2025, 36(11): 5134–5157 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/7322.htm [doi: 10.13328/j.cnki.jos.007322]

[6] Shang LY, Zhang Y, Zhang D, Wang D. FauxWard: A graph neural network approach to fauxtography detection using social media comments. Social Network Analysis and Mining, 2020, 10(1): 76. [doi: 10.1007/s13278-020-00689-w]

[7] Liu NH, Shin D, Hu X. Contextual outlier interpretation. In: Proc. of the 27th Int'l Joint Conf. on Artificial Intelligence. Stockholm: IJCAI, 2018. 2461–2467. [doi: 10.24963/ijcai.2018/341]

[8] Macha M, Akoglu L. Explaining anomalies in groups with characterizing subspace rules. Data Mining and Knowledge Discovery, 2018, 32(5): 1444–1480. [doi: 10.1007/s10618-018-0585-7]

[9] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature, 2015, 521(7553): 436–444. [doi: 10.1038/nature14539]

[10] Zhou J, Cui GQ, Hu SD, Zhang ZY, Yang C, Liu ZY, Wang LF, Li CC, Sun MS. Graph neural networks: A review of methods and applications. AI Open, 2020, 1: 57–81. [doi: 10.1016/j.aiopen.2021.01.001]

[11] McMahan B, Moore E, Ramage D, Hampson S, Arcas BAY. Communication-efficient learning of deep networks from decentralized data. In: Proc. of the 20th Int'l Conf. on Artificial Intelligence and Statistics. Fort Lauderdale: PMLR, 2017. 1273–1282.

[12] Zhang HD, Shen T, Wu F, Yin MY, Yang HX, Wu C. Federated graph learning—A position paper. arXiv:2105.11099, 2021.

[13] Ye M, Fang XW, Du B, Yuen PC, Tao DC. Heterogeneous federated learning: State-of-the-art and research challenges. ACM Computing Surveys, 2024, 56(3): 79. [doi: 10.1145/3625558]

[14] Li T, Sahu AK, Zaheer M, Sanjabi M, Talwalkar A, Smith V. Federated optimization in heterogeneous networks. In: Proc. of the 3rd Conf. on Machine Learning and Systems. Austin: MLSys, 2020. 429–450.

[15] Li QB, He BS, Song D. Model-contrastive federated learning. In: Proc. of the 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 10708–10717. [doi: 10.1109/CVPR46437.2021.01057]

[16] Wan GC, Huang WK, Ye M. Federated graph learning under domain shift with generalizable prototypes. In: Proc. of the 38th AAAI Conf. on Artificial Intelligence. Vancouver: AAAI Press, 2024. 15429–15437. [doi: 10.1609/aaai.v38i14.29468]

[17] Zhang K, Yang C, Li XX, Sun LC, Yiu SM. Subgraph federated learning with missing neighbor generation. In: Proc. of the 35th Int'l Conf. on Neural Information Processing Systems. Curran Associates Inc., 2021. 511.

[18] Zhang K, Sun LC, Ding BL, Yiu SM, Yang C. Deep efficient private neighbor generation for subgraph federated learning. In: Proc. of 2024 SIAM Int'l Conf. on Data Mining (SDM). Houston: Society for Industrial and Applied Mathematics, 2024. 806–814. [doi: 10.1137/1.9781611978032.92]

[19] Li T, Hu SY, Beirami A, Smith V. Ditto: Fair and robust federated learning through personalization. In: Proc. of the 38th Int'l Conf. on Machine Learning. PMLR, 2021. 6357–6368.

[20] Gao YJ, Wang PF, Liu L, Ma HD. Personalized federated learning method based on attention-enhanced meta-learning network. Journal of Computer Research and Development, 2024, 61(1): 196–208. (in Chinese with English abstract). [doi: 10.7544/issn1000-1239.202220922]

[21] Fan ZN, Fang H, Zhou ZR, Pei J, Friedlander MP, Liu CX, Zhang Y. Improving fairness for data valuation in horizontal federated learning. In: Proc. of the 38th IEEE Int'l Conf. on Data Engineering (ICDE). Kuala Lumpur: IEEE, 2022. 2440–2453. [doi: 10.1109/ICDE53745.2022.00228]

[22] Zhao J, Zhu XH, Wang JZ, Xiao J. Efficient client contribution evaluation for horizontal federated learning. In: Proc. of the 2021 IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP). Toronto: IEEE, 2021. 3060–3064. [doi: 10.1109/ICASSP39728.2021.9413377]

[23] Lv HT, Zheng ZZ, Luo T, Wu F, Tang SJ, Hua LF, Jia RF, Lv CF. Data-free evaluation of user contributions in federated learning. In: Proc. of the 19th Int'l Symp. on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt). Philadelphia: IEEE, 2021. 1–8. [doi: 10.23919/WiOpt52861.2021.9589136]

[24] Xu XY, Lyu LJ, Ma XJ, Miao CL, Foo CS, Low BKH. Gradient driven rewards to guarantee fairness in collaborative machine learning. In: Proc. of the 35th Int'l Conf. on Neural Information Processing Systems. Curran Associates Inc., 2021. 1232.

[25] Pan CL, Xu JR, Yu Y, Yang ZQ, Wu QB, Wang CB, Chen L, Yang Y. Towards fair graph federated learning via incentive mechanisms. In: Proc. of the 38th AAAI Conf. on Artificial Intelligence. Vancouver: AAAI Press, 2024. 14499–14507. [doi: 10.1609/aaai.v38i13.29365]

[26] Tan ZH, Wan GC, Huang WK, Ye M. FedSSP: Federated graph learning with spectral knowledge and personalized preference. In: Proc. of the 38th Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2024. 1090.

[27] Zehtabi S, Hosseinalipour S, Brinton CG. Decentralized event-triggered federated learning with heterogeneous communication thresholds. In: Proc. of the 61st IEEE Conf. on Decision and Control (CDC). Cancun: IEEE, 2022. 4680–4687. [doi: 10.1109/CDC51059.

2022.9993258]

[28] Suzumura T, Zhou Y, Baracaldo N, Ye GN, Houck K, Kawahara R, Anwar A, Stavarache LL, Watanabe W, Loyola P, Klyashtorny D, Ludwig H, Bhaskaran K. Towards federated graph learning for collaborative financial crimes detection. arXiv:1909.12946, 2019.

[29] Pan ZR, Wang GZ, Li ZN, Chen LF, Bian Y, Lai ZY. 2SFGL: A simple and robust protocol for graph-based fraud detection. In: Proc. of the 2022 IEEE Int'l Conf. on Cloud Computing Technology and Science (CloudCom). Bangkok: IEEE, 2022. 194–201. [doi: 10.1109/CloudCom55334.2022.00036]

[30] Li MQ, Walsh J. FedGAT-DCNN: Advanced credit card fraud detection using federated learning, graph attention networks, and dilated convolutions. Electronics, 2024, 13(16): 3169. [doi: 10.3390/electronics13163169]

[31] Guan ZL, Du JP, Xue Z, Wang PW, Pan ZH, Wang XY. Personalized public safety event detection based on reinforcement federated GNN. Ruan Jian Xue Bao/Journal of Software, 2024, 35(4): 1774–1789 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/7019.htm [doi: 10.13328/j.cnki.jos.007019]

[32] Yang YJ, Wang L, Wang YH. Rumor detection based on source information and gating graph neural network. Journal of Computer Research and Development, 2021, 58(7): 1412–1424. (in Chinese with English abstract). [doi: 10.7544/issn1000-1239.2021.20200801]

[33] Chen B, Zhang J, Zhang XK, Dong YX, Song J, Zhang P, Xu KB, Kharlamov E, Tang J. GCCAD: Graph contrastive coding for anomaly detection. IEEE Trans. on Knowledge and Data Engineering, 2023, 35(8): 8037–8051. [doi: 10.1109/TKDE.2022.3200459]

[34] Chen JY, Zhu GH, Yuan CF, Huang YH. Boosting graph anomaly detection with adaptive message passing. In: Proc. of the 12th Int'l Conf. on Learning Representations. Vienna: OpenReview.net, 2024. 1–22.

[35] Wang CX, Wang KY, Wang MQ. Malicious node detection based on semi-supervised and self-supervised graph representation learning. Ruan Jian Xue Bao/Journal of Software, 2025, 36(5): 2288–2307 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/7211.htm [doi: 10.13328/j.cnki.jos.007211]

[36] Gao Y, Wang X, He XN, Liu ZG, Feng HM, Zhang YD. Alleviating structural distribution shift in graph anomaly detection. In: Proc. of the 16th ACM Int'l Conf. on Web Search and Data Mining. Singapore: ACM, 2023. 357–365. [doi: 10.1145/3539597.3570377]

[37] Tan Y, Long GD, Liu L, Zhou TY, Lu QH, Jiang J, Zhang CQ. FedProto: Federated prototype learning across heterogeneous clients. In: Proc. of the 36th AAAI Conf. on Artificial Intelligence. AAAI Press, 2022. 8432–8440. [doi: 10.1609/aaai.v36i8.20819]

[38] Wang Y, Li GL, Li KY. Survey on contribution evaluation for federated learning. Ruan Jian Xue Bao/Journal of Software, 2023, 34(3): 1168–1192 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6786.htm [doi: 10.13328/j.cnki.jos.006786]

[39] Song TS, Tong YX, Wei SY. Profit allocation for federated learning. In: Proc. of the 2019 IEEE Int'l Conf. on Big Data (Big Data). Los Angeles: IEEE, 2019. 2577–2586. [doi: 10.1109/BigData47090.2019.9006327]

[40] Li XK, Wu ZY, Zhang WT, Zhu YL, Li RH, Wang GR. FedGTA: Topology-aware averaging for federated graph learning. Proc. of the VLDB Endowment, 2024, 17(1): 41–50. [doi: 10.14778/3617838.3617842]

[41] Huang WK, Wan GC, Ye M, Du B. Federated graph semantic and structural learning. In: Proc. of the 32nd Int'l Joint Conf. on Artificial Intelligence. Macao: IJCAI, 2024. 3830–3838. [doi: 10.24963/ijcai.2023/426]

[42] Chen C, Xu ZY, Hu WB, Zheng ZB, Zhang J. FedGL: Federated graph learning framework with global self-supervision. Information Sciences, 2024, 657: 119976. [doi: 10.1016/j.ins.2023.119976]

[43] Page L, Brin S, Motwani R, Winograd T. The PageRank citation ranking: Bringing order to the Web. In. Proc. of the 8th Int'l World Wide Web Conf. 1999.

[44] Tang JH, Li JJ, Gao ZQ, Li J. Rethinking graph neural networks for anomaly detection. In: Proc. of the 39th Int'l Conf. on Machine Learning. Baltimore: PMLR, 2022. 21076–21089.

[45] McAuley JJ, Leskovec J. From amateurs to connoisseurs: Modeling the evolution of user expertise through online reviews. In: Proc. of the 22nd Int'l Conf. on World Wide Web. Rio de Janeiro: ACM, 2013. 897–908. [doi: 10.1145/2488388.2488466]

[46] Wang Z, Kuang WR, Xie YX, Yao LY, Li YL, Ding BL, Zhou JR. FederatedScope-GNN: Towards a unified, comprehensive and efficient package for federated graph learning. In: Proc. of the 28th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining. Washington: ACM, 2022. 4110–4120. [doi: 10.1145/3534678.3539112]

[47] Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008, 2008(10): P10008. [doi: 10.1088/1742-5468/2008/10/P10008]

[48] Mu XT, Shen YL, Cheng K, Geng XL, Fu JX, Zhang T, Zhang ZW. FedProc: Prototypical contrastive federated learning on non-IID data. Future Generation Computer Systems, 2023, 143: 93–104. [doi: 10.1016/j.future.2023.01.019]

[49] Hamilton WL, Ying Z, Leskovec J. Inductive representation learning on large graphs. In: Proc. of the 31st Int'l Conf. on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 1025–1035.

附中文参考文献

[1] 宋凌云, 马卓源, 李战怀, 尚学群. 面向金融风险预测的时序图神经网络综述. 软件学报, 2024, 35(8): 3897–3922. http://www.jos.org.cn/1000-9825/7087.htm [doi: 10.13328/j.cnki.jos.007087]

[5] 殷茗, 乔胜, 陈威, 姜继娇. 基于群体情绪稳态化的社交网络谣言检测方法. 软件学报, 2025, 36(11): 5134–5157. http://www.jos.org.cn/1000-9825/7322.htm [doi: 10.13328/j.cnki.jos.007322]

[20] 高雨佳, 王鹏飞, 刘亮, 马华东. 基于注意力增强元学习网络的个性化联邦学习方法. 计算机研究与发展, 2024, 61(1): 196–208. [doi: 10.7544/issn1000-1239.202220922]

[31] 管泽礼, 杜军平, 薛哲, 王沛文, 潘圳辉, 王晓阳. 基于强化联邦 GNN 的个性化公共安全突发事件检测. 软件学报, 2024, 35(4): 1774–1789. http://www.jos.org.cn/1000-9825/7019.htm [doi: 10.13328/j.cnki.jos.007019]

[32] 杨延杰, 王莉, 王宇航. 融合源信息和门控图神经网络的谣言检测研究. 计算机研究与发展, 2021, 58(7): 1412–1424. [doi: 10.7544/issn1000-1239.2021.20200801]

[35] 王晨旭, 王凯月, 王梦勤. 基于半监督和自监督图表示学习的恶意节点检测. 软件学报, 2025, 36(5): 2288–2307. http://www.jos.org.cn/1000-9825/7211.htm [doi: 10.13328/j.cnki.jos.007211]

[38] 王勇, 李国良, 李开宇. 联邦学习贡献评估综述. 软件学报, 2023, 34(3): 1168–1192. http://www.jos.org.cn/1000-9825/6786.htm [doi: 10.13328/j.cnki.jos.006786]

附录 A

基于第 3.4 节的假设条件, 本附录提供定理 1 的详细证明过程.

定理 1. FedRPDA 的非凸收敛性. 基于上述假设, 在非凸条件假设下, FedRPDA 算法经过 T 轮通信后可以达到如下收敛保证.

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\|\nabla\mathcal{L}\left(w_g^t\right)\|^2\right]\leqslant\frac{4\beta^2}{\eta KT}\left(\mathcal{L}\left(w_g^0\right)-\mathcal{L}^*\right)+2\beta^2L^2\eta^2K\left(D^2+\sigma^2\right)+\frac{2\kappa^2}{K}+2\beta^2L\eta\sigma^2+4\beta^2\lambda L_2G \tag{A1}$$

证明: 考虑在通信轮次 t 时客户端 n 的第 k 步本地更新:

$$w_{n,k+1}^t=w_{n,k}^t-\eta\nabla\mathcal{L}_n\left(w_{n,k}^t;\xi_k\right) \tag{A2}$$

根据假设 1, 可以得到:

$$\mathcal{L}_n\left(w_{n,k+1}^t\right)\leqslant\mathcal{L}_n\left(w_{n,k}^t\right)+\left\langle\nabla\mathcal{L}_n\left(w_{n,k}^t\right),-\eta\nabla\mathcal{L}_n\left(w_{n,k}^t;\xi_k\right)\right\rangle+\frac{L}{2}\left\|\eta\nabla\mathcal{L}_n\left(w_{n,k}^t;\xi_k\right)\right\|^2 \tag{A3}$$

根据假设 2、假设 3 并取期望:

$$\mathbb{E}\left[\mathcal{L}_n\left(w_{n,k+1}^t\right)\right]\leqslant\mathcal{L}_n\left(w_{n,k}^t\right)-\eta\left\|\nabla\mathcal{L}_n\left(w_{n,k}^t\right)\right\|^2+\frac{L\eta^2}{2}\left(\left\|\nabla\mathcal{L}_n\left(w_{n,k}^t\right)\right\|^2+\sigma_n^2\right) \tag{A4}$$

在训练目标函数公式 (18) 中, 需要考虑多样化风险平均聚合项 $\mathcal{L}_p$ 的额外影响, 由假设 5 和中值定理, 存在常数 $G>0$ 使得:

$$\left\|\mathcal{L}_p\left(w_{n,k+1}^t\right)-\mathcal{L}_p\left(w_{n,k}^t\right)\right\|\leqslant G\eta\left\|\nabla\mathcal{L}_n\left(w_{n,k}^t;\xi_k\right)\right\| \tag{A5}$$

取期望并应用 Cauchy-Schwarz 不等式:

$$\mathbb{E}\left[\left\|\mathcal{L}_p\left(w_{n,k+1}^t\right)-\mathcal{L}_p\left(w_{n,k}^t\right)\right\|\right]\leqslant\eta G\mathbb{E}\left[\left\|\nabla\mathcal{L}_n\left(w_{n,k}^t\right)\right\|\right] \tag{A6}$$

损失项 $\mathcal{L}_n$ 的偏差满足:

$$\mathbb{E}\left[\mathcal{L}_n\left(w_{n,k+1}^t\right)\right]\leqslant\mathcal{L}_n\left(w_{n,k}^t\right)-\left(\eta-\frac{L\eta^2}{2}\right)\left\|\nabla\mathcal{L}_n\left(w_{n,k}^t\right)\right\|^2+\frac{L\eta^2}{2}\sigma_n^2+\lambda\eta G\mathbb{E}\left[\left\|\nabla\mathcal{L}_n\left(w_{n,k}^t\right)\right\|\right] \tag{A7}$$

应用 Young 不等式到末项, 整体化简得:

$$\mathbb{E}\left[\mathcal{L}_n\left(w_{n,K}^t\right)\right]\leqslant\mathcal{L}_n\left(w_g^t\right)-\sum_{k=0}^{K-1}\left(\eta-\frac{L\eta^2}{2}-\frac{\lambda}{2}\right)\left\|\nabla\mathcal{L}_n\left(w_{n,k}^t\right)\right\|^2+\frac{LK\eta^2}{2}\sigma_n^2+\lambda\eta L_2KG \tag{A8}$$

对于服务端加权聚合项:

$$\mathbb{E}\left[\mathcal{L}\left(w_g^{t+1}\right)\right] \leqslant \sum_{n=1}^{N} \rho_n^t \left(\mathcal{L}_n\left(w_g^t\right) - \frac{\eta}{2} \sum_{k=0}^{K-1} \left\| \nabla \mathcal{L}_n\left(w_{n,k}^t\right) \right\|^2 + \frac{LK\eta^2}{2} \sigma_n^2 + \lambda \eta L_2 K G \right) \tag{A9}$$

由假设 4, 存在常数 $\beta^2 \geqslant 1$ 与 $\kappa^2 \geqslant 0$, 使得:

$$\sum_{n=1}^{N} \rho_n^t \left\| \nabla \mathcal{L}_n\left(w_{n,k}^t\right) \right\|^2 \geqslant \frac{1}{\beta^2} \left\| \sum_{n=1}^{N} \rho_n \nabla \mathcal{L}_n\left(w_{n,k}^t\right) \right\|^2 - \kappa^2 \tag{A10}$$

分解全局梯度:

$$\left\| \sum_{n=1}^{N} \rho_n^t \nabla \mathcal{L}_n\left(w_{n,k}^t\right) \right\|^2 \geqslant \frac{1}{2} \left\| \nabla \mathcal{L}\left(w_g^t\right) \right\|^2 - L^2 \sum_{n=1}^{N} \rho_n^t \left\| w_{n,k}^t - w_g^t \right\|^2 \tag{A11}$$

由梯度有界性 $D \triangleq \max_{w,n} \|\nabla \mathcal{L}_n(w_n)\|$, 得到本地参数偏差界:

$$\mathbb{E}\left[\left\| w_{n,k}^t - w_g^t \right\|^2\right] \leqslant \eta^2 k \left(D^2 + \sigma^2\right) \tag{A12}$$

对 $t = 0$ 到 $T - 1$ 求和, 可以得到:

$$\frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\left[\left\| \nabla \mathcal{L}\left(w_g^t\right) \right\|^2\right] \leqslant \frac{4\beta^2}{\eta K T} \left(\mathcal{L}\left(w_g^0\right) - \mathcal{L}^*\right) + 2\beta^2 L^2 \eta^2 K \left(D^2 + \sigma^2\right) + \frac{2\kappa^2}{K} + 2\beta^2 L \eta \sigma^2 + 4\beta^2 \lambda L_2 G \tag{A13}$$

其中, $\mathcal{L}^* \triangleq \min_w \mathcal{L}(w)$.

作者简介

杨家震, 硕士生, 主要研究领域为人工智能安全, 联邦学习, 基于图的数据挖掘.

邱天, 博士生, 主要研究领域为机器学习, 表征学习, 基于图的数据挖掘.

陈可嘉, 博士生, 主要研究领域为可信的生成式人工智能, 高效深度学习.

段明江, 博士生, 主要研究领域为反欺诈, 基于图的数据挖掘.

蒋健, 高级工程师, 主要研究领域为新一代信息技术, 人工智能.

胡泽远, 硕士, 主要研究领域为反欺诈, 人工智能安全, 大数据分析.

宋明黎, 博士, 教授, 博士生导师, CCF 专业会员, 主要研究领域为计算机视觉, 模式识别与人工智能, 机器学习.

冯尊磊, 博士, 副教授, 博士生导师, CCF 专业会员, 主要研究领域为医学图像分析, 模型优化, 基于图的数据挖掘.