

DOI: 10.3901/JME.2023.21.283

区块链和边缘计算赋能的联邦学习故障诊断框架^{*}

邵海东¹ 肖一鸣¹ 闵志闪¹ 韩淞宇¹ 张海舟²

(1. 湖南大学机械与运载工程学院 长沙 410082;

2. 南京电子技术研究所 南京 210039)

摘要: 工业物联网助推机械故障诊断步入大数据时代, 然而因各节点需要共享本地的私有数据而造成隐私泄露是当前工业物联网亟需解决的问题。联邦学习有望应用于工业物联网以实现在私有数据不离开本地存储的前提下, 协同各节点训练诊断模型。然而, 联邦学习面临着以下诸多挑战。首先, 联邦学习的中心化架构极易引发单点故障。其次, 工业物联网中各节点的故障数据通常是非独立同分布的, 以致联邦学习难以收敛。再次, 联邦学习缺乏防御手段来阻止恶意节点的攻击。最后, 联邦学习需要激励机制来鼓励节点分享资源。针对这些挑战, 提出了一种区块链和边缘计算赋能的联邦学习故障诊断框架, 采用去中心化的模式保障工业物联网中机械设备故障数据的隐私和安全。在此框架中, 构造了一种特征对比损失函数来解决非独立同分布问题, 设计了一种拜占庭容错的评分机制来抵抗投毒攻击, 并开发了一种基于信誉的激励算法来评估应给予节点的奖励。所提方法被应用于工业物联网中风力发电机的行星齿轮箱故障诊断模拟场景, 在私有本地数据不泄露的前提下, 展现出最优的综合性能。

关键词: 区块链; 边缘计算; 故障诊断; 联邦学习; 工业物联网

中图分类号: TH17

Blockchain and Edge Computing Enabled Federated Learning Fault Diagnosis Framework

SHAO Haidong¹ XIAO Yiming¹ MIN Zhishan¹ HAN Songyu¹ ZHANG Haizhou²

(1. College of Mechanical and Vehicle Engineering, Hunan University, Changsha 410082;

2. Nanjing Research Institute of Electronics Technology, Nanjing 210039)

Abstract: The industrial Internet of Things (IIoT) promotes mechanical fault diagnosis into the era of big data, however, privacy leakage caused by the need to share local private data among IIoT nodes is an urgent problem to be solved. Federated learning (FL) is expected to be applied to IIoT, which enables nodes to collaboratively train diagnostic models without making private data leave local storage. However, there are many challenges faced by the FL. Firstly, the centralized architecture of FL is highly susceptible to single point of failure. Moreover, the fault data of nodes in the IIoT are usually not independent and identically distributed (non-IID), which makes it difficult for the FL to converge. In addition, the FL lacks defense measures to prevent attacks conducted by malicious nodes. Finally, the FL needs incentive mechanisms to encourage nodes to share resources. Aiming at the challenges introduced above, a blockchain and edge computing enabled FL fault diagnosis framework is proposed, which adopts a decentralized mode to ensure the privacy and security of mechanical equipment fault data in the IIoT. In the proposed framework, a feature-contrastive loss function is constructed to address the non-IID problem. A Byzantine-tolerance scoring mechanism is designed to resist poisonous attacks. A reputation-based incentive algorithm is developed to evaluate the rewards owed to nodes. The proposed method is applied to a simulation scenario of planetary gearbox fault diagnosis for wind turbines in the IIoT, demonstrating its optimal overall performance without the disclosure of local private data.

Key words: blockchain; edge computing; fault diagnosis; federated learning; industrial Internet of Things

^{*} 国家自然科学基金(52275104)和湖南省自然科学基金优秀青年科学基金(2021JJ20017)资助项目。20221226 收到初稿, 20230701 收到修改稿

0 前言

随着“中国智能制造 2025”，“工业 4.0”等概念的不断推广，工业物联网技术受到了全球研究人员更广泛的关注并取得了长足的发展，推动了机械故障诊断步入大数据时代^[1-2]。工业物联网为智能故障诊断技术提供了大量机械设备健康监测数据，可有效帮助其挖掘隐藏在故障数据背后的潜在信息和价值。然而，传统的诊断方法需要工业物联网中的各节点共享本地的原始故障数据来协同训练诊断模型。考虑到在现代工业中，数据隐私至关重要，私有数据离开本地存储通常是被严格禁止的^[3]，因此这种方法可行性较低。

联邦学习是一种分布式机器学习模式，可用于解决用户私有故障数据的隐私问题^[4]。在联邦学习中，各用户无须上传本地数据，而是使用本地数据训练一个局部模型，并将模型权重上传至中心服务器来融合一个具备所有用户的诊断知识的全局模型，从而在私有数据未泄露的前提下，完成诊断模型的协同训练。然而，应用于工业物联网的联邦学习故障诊断仍面临诸多挑战。①目前工业物联网中的联邦学习系统大多采用以中心服务器为核心的中心化架构，极易引发单点故障^[5]。②工业物联网中不同节点的故障数据通常是非独立同分布的，即部分节点会存在缺少某些故障类别的数据，或是数据样本不平衡的情况。这会造成在联邦学习过程中一方所训练的模型的局部优化目标远离全局优化目标，从而阻碍全局模型收敛至最优^[6]。③现有的联邦学习缺乏有效的检测机制来确保协同训练的可靠性。例如，工业物联网中的恶意节点可能制造毒性样本来污染局部模型，从而扰乱全局模型的训练方向。此外，在非独立同分布条件下，由良性节点所训练的局部模型可能会同样偏离全局模型，使得恶意节点的检测极具挑战性^[7]。④考虑到联邦学习的性能极度依赖于足够数量的良性节点的参与，因此有必要设计一个激励算法来鼓励更多的良性节点分享他们的计算和通信资源，并惩罚恶意节点^[8]。

作为一种去中心化架构，区块链有望消除联邦学习需要中心服务器来管理多方数据传输的依赖性，实现去中心化的联邦学习^[5]。在区块链赋能的联邦学习中，联邦学习的模型参数聚合操作可交由被选中的矿工来执行，从而避免了单点故障问题。此外，区块链的共识协议完美地解决了去中心化架构中互不信任的多个参与方如何达成共识的难题。

最后，区块链的激励机制也能鼓励更多的节点加入联邦学习。考虑到边缘计算能提供低延迟的服务且通常配备了充足的通信和计算资源，提出了一种区块链和边缘计算赋能的联邦学习框架来实现工业物联网下的机械设备故障诊断。主要创新和贡献如下。

(1) 搭建了一种新颖的框架来使联邦学习具备区块链的多种特性，进一步维护工业物联网中各节点的数据隐私和安全。所提方法被应用于模拟的工业物联网中风力发电机的行星齿轮箱故障诊断场景。实验结果表明所提方法优于多种流行的联邦学习方法。

(2) 构造了一种特征对比损失函数来最小化局部模型和全局模型所学表征之间的差异，在非独立同分布条件下训练无偏的全局模型。

(3) 设计了一种拜占庭容错的评分机制以在非独立同分布条件下减小恶意节点对全局模型训练的负面影响。

(4) 开发了一种基于信誉的激励算法来评估应给予各节点的奖励或惩罚。

1 相关工作

近年来，国内外学者已陆续开展了关于联邦学习，区块链和边缘计算等技术的交叉研究。由于在现有文献中，探讨如何解决非独立同分布问题，如何设计检测机制，以及如何开发激励算法的这三类工作与所提方法密切相关，因此将重点评述这三类研究以体现所提方法的优势。

(1) 非独立同分布问题的解决：2018 年，ZHAO 等^[9]通过全局共享一个标准训练数据集来减轻非独立同分布问题对联邦学习性能的影响。尽管这种方法显著提升了联邦学习的准确率，但其存在需要事先收集一个高质量共享数据集的局限性。2020 年，KARIMIREDDY 等^[10]提出了一种 SCAFFOLD 算法来限制局部模型和全局模型更新方向之间的漂移。2021 年，ZHANG 等^[11]利用各节点的正负类数据质心间的距离开发了一种 CDW_FedAvg 算法来解决数据异质性问题。然而，SCAFFOLD 算法所需要的额外控制变量使其通信成本增加了一倍，CDW_FedAvg 算法也局限于二分类场景。针对以上问题，所构造的特征对比损失函数不仅消除了对共享训练数据集的需求，而且不增加额外的通信成本，同时也能适用于各种分类场景。

(2) 检测机制设计：2021 年，LI 等^[12]提出了一种基于区块链的共识委员会联邦学习模型，允许部

分节点使用本地数据集来验证其他节点的局部模型。2021 年, XU 等^[13]开发了一种基于准确率的恶意节点检测机制, 可使系统内的验证支持者使用标准数据集来评估计算支持者的模型质量。类似地, LIU 等^[14]在 2021 年所设计的联邦学习框架中的中央聚合器也可使用测试数据集来筛选有利于全局模型的局部更新。然而, 上述利用数据集进行验证的方法在训练早期效果较差。更重要的是, 文献中的方法必须要在联邦学习系统中预置一个标准的验证数据集^[13-14]。2021 年, ZHAO 等^[8]采用 Multi-KRUM 算法来从事务池中搜索和删除恶意权重, 放松了对验证数据集的需求。然而, 在非独立同分布条件下, 上述所有方法都可能导致有部分类别数据缺失的良性节点被视作恶意节点。幸运的是, 所设计的拜占庭容错的评分机制能较好地解决此类难题。

(3) 激励算法开发: 区块链通过代币奖励来吸引参与者, 奖励的大小取决于区块链对参与者的评价角度。2019 年, KANG 等^[15]设计了一种以节点的本地训练时间为指标的评估方法。2019 年, FENG 等^[16]认为奖励大小应与节点所提供的本地数据集的大小成正比。显然, 上述方法的评价维度过于单一, 不能充分衡量各节点对联邦学习系统的贡献。2021 年, ZHANG 等^[11]从本地数据集质量的角度来评价各节点, 但该质量是通过节点的正负类数据质心间的距离来决定的, 使得此方法无法泛用至多分

类场景。不同的是, 所开发的基于信誉的激励算法通过丰富的维度计算节点的信誉来充分评价节点在过去训练中的表现, 有效避免了上述问题。

2 所提诊断框架的整体结构

如图 1 所示, 所提的区块链和边缘计算赋能的联邦学习故障诊断框架包含 K 个客户端, 多个边缘计算服务器以及作为平台的去中心化应用程序。客户端可以是工业物联网中拥有有限的数据集和计算资源的小型企业或制造商等, 他们希望在保护本地故障数据隐私的前提下与其他客户端协作训练一个更优秀的诊断模型。考虑到边缘计算服务器通常配备了充足的计算和通信资源, 他们在此框架中作为矿工来提供区块链的验证, 共识等其他服务, 并赚取相应的代币作为收入。去中心化应用程序是一种部署在区块链上基于智能合约自动执行的新型应用程序。客户端会关联至与自己地理位置最近的边缘计算服务器来减少通信延迟。在任何节点加入所提框架之前, 他们必须向平台提交注册申请以验明身份, 注册信息应包含其本地数据集的大小 $n^k (1 \leq k \leq K)$, 然后平台将为节点分配用于签名的公钥和私钥。由于所提框架中已不存在中心服务器, 因此任何节点都可作为模型需求者向平台提交协作训练请求。所提框架的一次完整训练流程可概括如下:

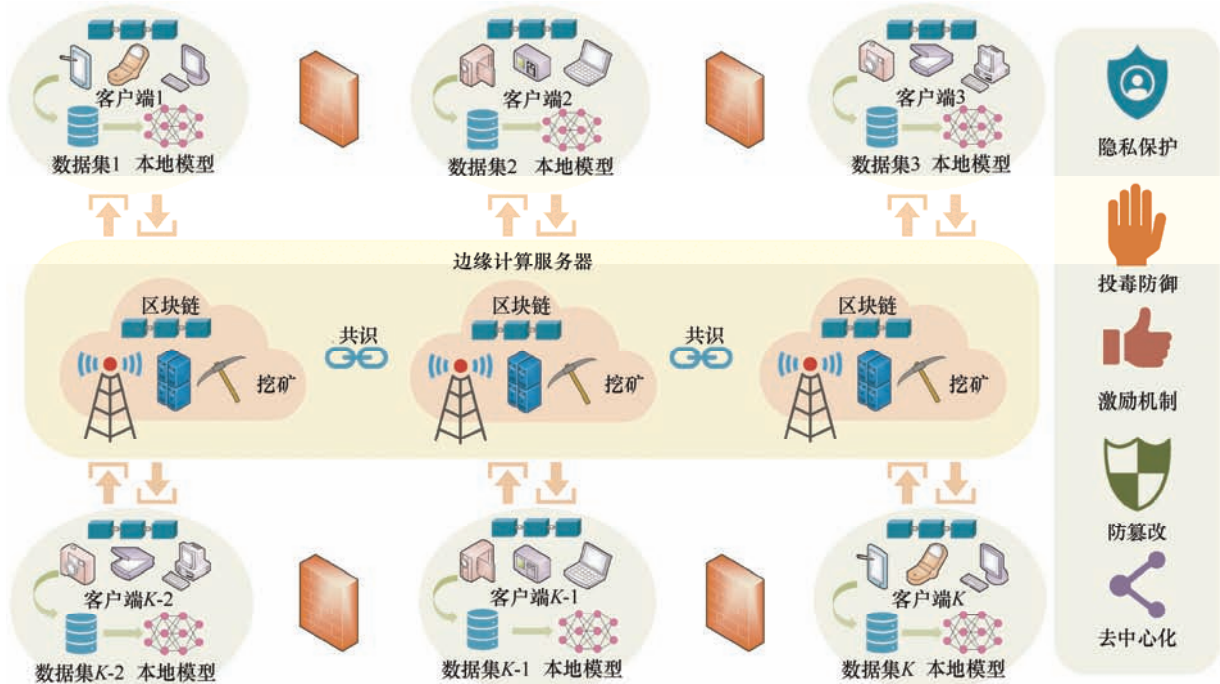


图 1 区块链和边缘计算赋能的联邦学习框架的结构示意图

步骤 1: 模型需求者向平台提交协作训练请求并提供相关信息, 包括: ① 诊断模型的目的; ② 故

障样本的要求和示例; ③ 诊断模型的结构及初始权重 w_0 , 全局训练总轮数 T ; ④ 愿意向参与者支付

的总代币数；⑤ 基于信誉的激励算法的参数设置。随后，平台将创建并发布包含上述信息的创世块以完成模型初始化。

步骤 2：平台挑选一组信誉良好的客户端(详见第 3.3 节)来参与本轮训练。所挑选的客户端首先使用本地数据和特征对比损失函数来完成本地模型更新，然后将更新后的本地模型的权重连同数字签名一起以区块链交易的形式发送至所关联的矿工。

步骤 3：在收到数据后，矿工需要验证签名的合法性来阻止攻击者篡改数据。具有合法签名的交易会被放入交易池等待权重评分。

步骤 4：平台通过共识协议挑选一组矿工组建验证委员会，其中具有最高优先级的委员会成员会被选举为拥有记账权的领导者(详见第 3.4 节)。

步骤 5：验证委员会执行拜占庭容错的评分机制为交易池内的所有权重评分并签名(详见第 3.2 节)。各客户端的信誉将根据委员会的评估结果进行更新(详见第 3.3 节)。领导者计算全局模型的权重并生成新的区块(详见第 3.2 节)。

步骤 6：验证委员会核查新区块的合法性并将经过核查的区块通过 Gossip 协议广播至全网，以在区块链中达成共识(详见第 3.4 节)。委员会成员及领导者随后会收到来自平台的挖矿奖励。

步骤 7：客户端下载新区块以获取全局权重来更新本地模型，并从步骤(2)开始进行下一轮的全局训练，直到达到全局训练总轮数。作为对客户保持良好训练行为以及持续提供有效更新的激励，训练奖励将在训练终止后基于客户端的最终声誉进行发放(详见第 3.3 节)。

3 所提诊断框架的关键组件

3.1 特征对比损失函数的构造

非独立同分布问题是联邦学习中的一个关键挑战。由于本地数据呈现出倾斜的数据分布，在全局模型基础上训练的局部模型甚至会比该全局模型学习到更差的表征^[17]。因此，构造了一个特征对比损失函数来限制局部模型在本地训练阶段的漂移。一方面，局部模型与全局模型所学表征间的差异将被减小。另一方面，相邻两轮的局部模型所学表征间的差异将被放大。

如图 2 所示，所训练的卷积神经网络(Convolutional neural network, CNN)模型由三部分组成，包括主干网络，瓶颈层以及输出层。特征对比损失函数由一个基于交叉熵的监督学习损失项 l_{sup} 和一个对比损失项 l_{con} 构成。为便于对所提方法

进行阐述，当模型权重为 w 时，使用 $R_w(\cdot)$ 来表示输出层之前的模型。

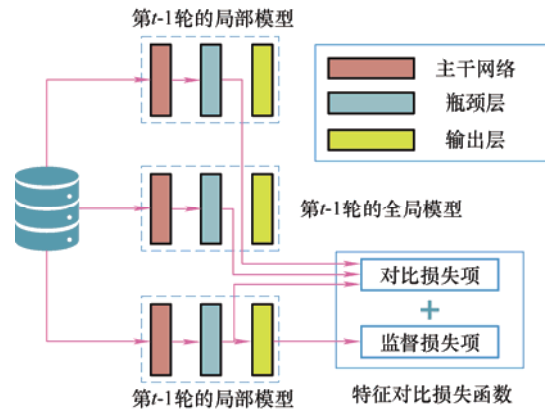


图2 特征对比损失函数的构造

在第 t 轮全局训练中，假设第 k 个客户端 C^k 收到了上一轮全局模型的权重 w_{t-1} 并尝试将其更新为 w_t^k 。令 $\mathcal{D}^k = \{x_i^k, y_i^k\}_{i=1}^{n^k}$ 为 C^k 的本地数据集，其中 x_i^k 是标签为 y_i^k 的第 i 个样本。为构造特征对比损失函数，首先需要获取三组特征表示：全局模型 w_{t-1} 所学的特征 $R_{w_{t-1}}(x_i^k)$ ；正在更新的局部模型 w_t^k 所学的特征 $R_{w_t^k}(x_i^k)$ ；以及上一轮更新完毕的局部模型 w_{t-1}^k 所学的特征 $R_{w_{t-1}^k}(x_i^k)$ 。

最大均值差异(Maximum mean discrepancy, MMD)是一种常用的用于衡量两组数据分布间差异的度量指标。MMD 的核心思想是两个相同的数据分布的各项统计指标也应该是相同的。具体来说，MMD 的定义为

$$\hat{d}_{\mathcal{H}}(p, q) = \left\| \mathbb{E}_p[\phi(X^s)] - \mathbb{E}_q[\phi(X^t)] \right\|_{\mathcal{H}}^2 \quad (1)$$

式中， \mathcal{H} 为再生核希尔伯特空间， p 和 q 为两个数据分布， X^s 和 X^t 分别是服从 p 和 q 的两个样本集， $\phi(\cdot)$ 表示将原始样本映射至再生核希尔伯特空间。因此， $R_{w_{t-1}}(x_i^k)$ 和 $R_{w_t^k}(x_i^k)$ 间的差异可通过 MMD 来计算。在实际中，MMD 可使用 $R_{w_{t-1}}(x_i^k)$ 和 $R_{w_t^k}(x_i^k)$ 的经验核均值嵌入间的平方距离来进行估计

$$\begin{aligned} l_{\text{mmd}}(w_t^k; w_{t-1}; x_i^k) &= \left\| \frac{1}{n^k} \sum_{i=1}^{n^k} \phi(R_{w_{t-1}}(x_i^k)) - \frac{1}{n^k} \sum_{j=1}^{n^k} \phi(R_{w_t^k}(x_j^k)) \right\|_{\mathcal{H}}^2 \\ &= \frac{1}{(n^k)^2} \sum_{i=1}^{n^k} \sum_{j=1}^{n^k} k(R_{w_{t-1}}(x_i^k), R_{w_{t-1}}(x_j^k)) + \\ &\quad \frac{1}{(n^k)^2} \sum_{i=1}^{n^k} \sum_{j=1}^{n^k} k(R_{w_t^k}(x_i^k), R_{w_t^k}(x_j^k)) - \\ &\quad \frac{2}{(n^k)^2} \sum_{i=1}^{n^k} \sum_{j=1}^{n^k} k(R_{w_{t-1}}(x_i^k), R_{w_t^k}(x_j^k)) \end{aligned} \quad (2)$$

式中, $k(\cdot)$ 为高斯核函数。相似地, $R_{w_{t-1}^k}(x_i^k)$ 和 $R_{w_t^k}(x_i^k)$ 间的差异也可通过此方法计算得到。

$$l_{con}(w_t^k; w_{t-1}^k; w_{t-1}^k; x_i^k) = l_{mmd}(w_t^k; w_{t-1}^k; x_i^k) - l_{mmd}(w_{t-1}^k; w_t^k; x_i^k) + \varepsilon \quad (3)$$

式中, ε 为一个边缘常数, 用于避免模型直接令 $l_{mmd}(w_t^k; w_{t-1}^k, x_i^k) = l_{mmd}(w_{t-1}^k, w_t^k, x_i^k)$ 而走捷径。

因此, 特征对比损失函数可被定义为

$$l = l_{sup}(w_t^k; (x_i^k, y_i^k)) + \mu l_{con}(w_t^k; w_{t-1}^k; w_{t-1}^k; x_i^k) \quad (4)$$

式中, μ 为一个正则化常数。

本地模型更新的优化目标为

$$\min_{w_t^k} \mathbb{E}_{(x_i^k, y_i^k) \sim D^k} [l_{sup}(w_t^k; (x_i^k, y_i^k)) + \mu l_{con}(w_t^k; w_{t-1}^k; w_{t-1}^k; x_i^k)] \quad (5)$$

3.2 拜占庭容错的评分机制的设计

联邦学习需要防御措施来抵抗来自工业物联网中的恶意节点的投毒攻击。如图 3a 所示, 在独立同分布条件下, 恶意的局部权重会倾向于偏离全局权重, 而良性的局部权重通常会与全局权重较为接近。因此, 可以考虑移除与全局权重有较大距离的局部权重来消除恶意客户端对联邦学习系统的影响。然而, 在非独立同分布条件下, 一个具有挑战性的问题被观察到了。如图 3b 所示, 良性客户端可能会因为其缺失某些类别的数据而导致其训练的局部权重同样远离全局权重。此外, 如图 3c 所示, 若多个有类别缺失的客户端的本地数据集可以相互补充, 则他们的权重之和的方向会与全局权重的方向趋于一致而协作为全局模型的收敛做出贡献。因此, 在非独立同分布条件下, 不能简单地认为远离全局权重的局部权重是无效的。

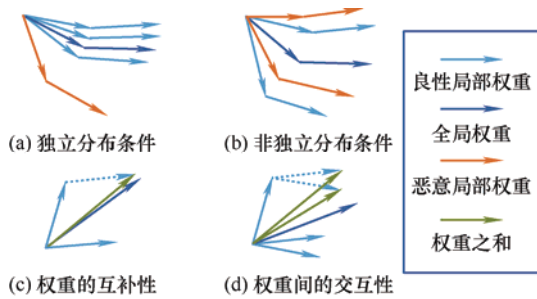


图 3 权重的矢量简图

根据以上发现, 设计了一种拜占庭容错的评分机制来为各局部权重设置一个置信度, 而不是直接丢弃异常权重。假设在交易池中共有 R 个权重, 该机制首先会评估各权重与其他权重的交互性。具体来说, 如图 3d 所示, 权重 w_t^k 与权重 $w_t^{k'}$ 的交互性可定义为这两个权重的加权平均值与全局权重 w_{t-1} 之间的余弦相似度。因此, 权重 w_t^k 与交易池中其他所

与三元组损失函数的形式类似, 对比损失项的定义式如下所示

有权重的交互性可量化为

$$\phi_t^k = \sum_{k'=1}^R \text{sim} \left(\frac{n^k}{n^k + n^{k'}} w_t^k + \frac{n^{k'}}{n^k + n^{k'}} w_t^{k'}, w_{t-1} \right) \quad (6)$$

式中, $\text{sim}(\cdot, \cdot)$ 为余弦相似度函数。然后, 机制会将所计算的所有交互性的值从大到小排序, 而排名在前 f 的权重会被直接视作可信权重, 其得分为 1。值得注意的是 f 的值应由模型提供者来设定。至此, 一个良性基线可利用这些可信权重计算得到

$$\hat{w}_t = \sum_{k=1}^f \frac{n^{\hat{k}}}{\hat{n}} w_t^{\hat{k}} \quad \hat{n} = \sum_{k=1}^f n^{\hat{k}} \quad (7)$$

式中, $w_t^{\hat{k}}$ 是第 \hat{k} 个可信权重。随后, 机制将计算剩余 $R-f$ 个权重与此基线之间的余弦相似度并将它们从大到小排序。则此 $R-f$ 个权重的得分可计算为

$$\eta_t^{\bar{k}} = 1 - r_t^{\bar{k}} / (R - f) \quad (8)$$

式中, $r_t^{\bar{k}}$ 是剩余 $R-f$ 个权重中的第 \bar{k} 个权重的排名。

该机制由验证委员会执行, 领导者需要使用所得到的评分计算本轮的全局权重

$$w_t = \sum_{k=1}^R \frac{n^k}{n} \cdot \eta_t^k \cdot w_t^k \quad n = \sum_{k=1}^R n^k \quad (9)$$

3.3 基于信誉的激励算法的开发

所开发的基于信誉的激励算法采用社交准则^[18]来约束各客户端的行为并基于他们的信誉为其发放训练奖励。具体地, 在训练初始化时, 第 k 个客户端会被分配一个取自集合 $\{0, 1, 2, \dots, \theta^{\max}\}$ 的整数 θ_0^k 作为其初始信誉, 其中 θ^{\max} 为信誉 θ 的最大值, 以及用于累计其被警告次数和判断其是否需要隔离的描述符 $\text{warning}(k)$ 和 $\text{isolation}(k)$ 。模型需求者应设置一个初始的社交阈值 h_0 并使其以每轮提高 β 的速度增长。需要注意的是 h_0 和 β 的值不应太大或太小, 否则都将使此社交准则失效。在随后的每轮训练中, 客户端的信誉会根据验证委员会对其权重的评分及当前的社交阈值进行调整。此算法通过智能合约自动执行, 所有客户端每轮的信誉值将被记录在区块链中以供审计回顾。算法的运行流程如图 4 所示。

在此规则下, 对于具有良好信誉即信誉不小于社交阈值的客户端, 其信誉将持续增长且所增加的值等于其权重得分。然而, 对于信誉跌至社交阈值之下的客户端, 算法将给予其一次调整机会作为警告。若该客户端的信誉持续低下, 则会被平台隔离。

被隔离的客户端的信誉将直接清零以作为惩罚。在隔离期间,客户端会被禁止参加训练直至其信誉以每轮加 1 的速度回升至社交阈值。因此,客户端的最终信誉与最终社交阈值的差值反映了其在过去训练中的社交状态,训练结束后发放给客户端的代币奖励可计算为

$$Token(k) = \max(0, \alpha \times (\theta_T^k - h_T)) \quad (10)$$

式中, α 为一个比例系数。

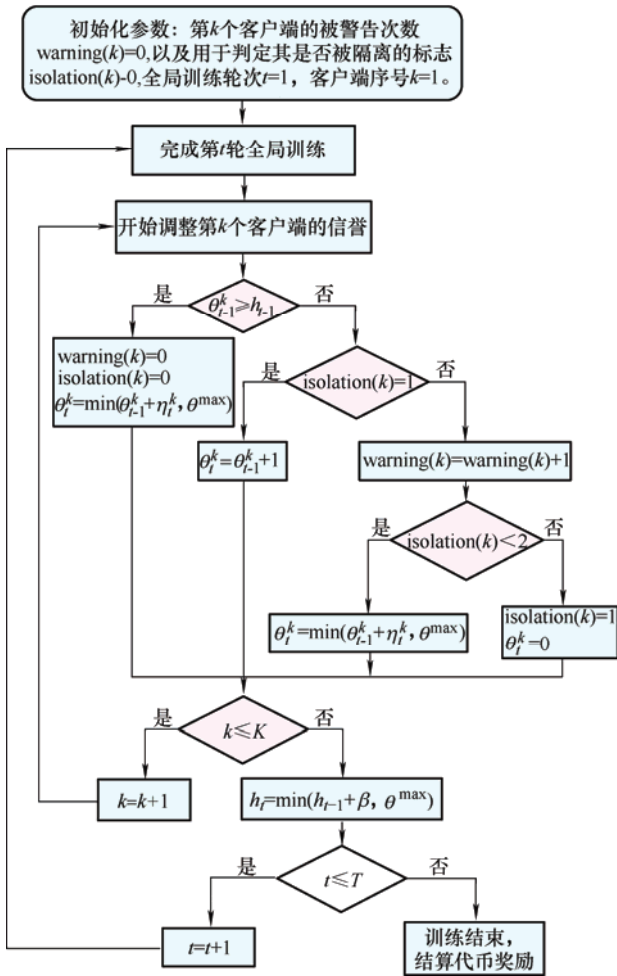


图4 基于信誉的激励算法运行流程图

3.4 共识协议

共识协议是区块链技术的核心,工作量证明是最常用的共识协议。然而,工作量证明十分浪费计算资源且其交易效率低下。Algorand 是一种基于权益证明和拜占庭容错的共识协议^[19],它能实现快速交易验证并能有效抵抗女巫攻击,DDoS 攻击等。因此,采用 Algorand 使区块链的各矿工达成共识。

Algorand 基于矿工的权益即其账户中的代币,使用可验证随机函数来从矿工中挑选领导者候选人组建验证委员会,并根据他们的优先级决定领导者。为通过矿工的权益选择矿工,矿工所拥有的每个代

币都应被视为一个子矿工。令 τ 为系统希望选择的子矿工的数量, M 为所有矿工的代币之和,则每个子矿工被选中的概率为 τ/M 。对于一个拥有 m 个代币的矿工,他首先会将其私钥输入至可验证随机函数来获取一个伪随机的哈希散列值 $hash$ 和一个证明 $proof$,然后将区间 $[0,1)$ 切分为 m 个子区间。若 $hash/2^{hashlen}$ 满足

$$hash/2^{hashlen} \in \left[\sum_{b=0}^B \binom{m}{b} p^b (1-p)^{m-b}, \sum_{b=0}^{B+1} \binom{m}{b} p^b (1-p)^{m-b} \right] \quad (11)$$

式中, $hashlen$ 为 $hash$ 的长度,则意味着矿工拥有 B 个被选中的子矿工。被选中的子矿工的数量决定了矿工的优先级,具有最高优先级的矿工会成为领导者。其他矿工可使用 $proof$ 来验证该矿工的确拥有 B 个被选中的子矿工。

领导者负责计算当前轮次的全局权重并生成新的区块。当超过 2/3 的委员会成员认可新区块时,此新区块将生效并链接到前一个区块,组成如图 5 所示的区块链。

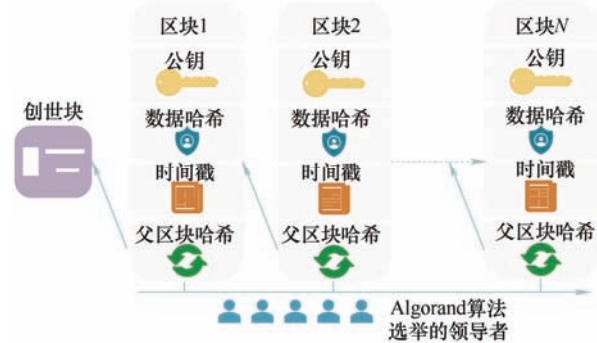


图5 基于 Algorand 算法的区块链结构

4 实验细节和结果分析

风力发电机被广泛部署于高山,海岛,荒野等风口处,常年经受极端工作条件的影响,从而可能会出现不同的故障模式,造成经济损失或安全事故。在大量风力发电机所形成的工业物联网架构下,各风力发电厂能相互传输风机装备健康监测数据,从而为风机装备设计更完善的故障预测和健康管理方案^[20]。然而,考虑到数据隐私问题,私有数据往往是被禁止离开本地存储的。因此,将所提方法应用于两个工业物联网中风力发电机的行星齿轮箱故障诊断模拟案例来验证所提方法的有效性。在两个案例中,客户端是配备了同种型号的风力发电机的风力发电厂,他们希望通过协作训练一个用于检测其关键部件行星齿轮箱的故障的智能诊断模型来保障

风力发电机的稳健运行。

4.1 数据集介绍

案例 1 所使用的数据集是来自西安交通大学的行星齿轮箱故障数据集^[21]，其实验平台如图 6a 所示，电机转速为 1 800 r/min，在行星齿轮箱的 x 和 y 方向分别部署了两个加速度传感器来采集振动信号，采样频率为 20 480 Hz。选用 x 方向的信号构造样本集，其详细信息列于表 1。

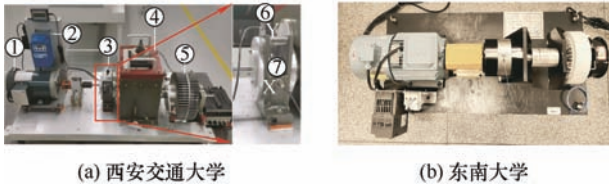


图 6 行星齿轮箱故障模拟试验台

表 1 西安交通大学行星齿轮箱故障数据集

健康状态	标签	训练/测试样本数	样本维度
正常	0		
轴承滚珠故障	1		
轴承内圈故障	2	960/40	
轴承外圈故障	3	(训练样本会被随机分配至框架内的所有客户端，测试样本用于测试全局模型的准确率)	1 000
轴承三种混合故障	4		
行星轮断齿	5		
行星轮缺齿	6		
行星轮齿根裂纹	7		
行星轮齿面磨损	8		

案例 2 所使用的数据集是来自东南大学的行星齿轮箱故障数据集，其实验平台如图 6b 所示，电机转速为 2 500 r/min，所部署的传感器为三通道加速

度传感器，采样频率为 12 000 Hz。选用第一个通道的信号构造样本集，其详细信息列于表 2。

表 2 东南大学行星齿轮箱故障数据集

健康状态	标签	训练/测试样本数	样本维度
正常	0		
齿圈断齿	1		
齿圈裂纹	2		
齿圈缺齿	3		
齿圈点蚀	4	768/32	
行星轮断齿	5	(训练样本会被随机分配至框架内的所有客户端，测试样本用于测试全局模型的准确率)	1 000
行星轮裂纹	6		
行星轮缺齿	7		
行星轮点蚀	8		
太阳轮断齿	9		
太阳轮裂纹	10		
太阳轮缺齿	11		
太阳轮点蚀	12		

所构造的样本在输入至模型前需经过快速傅里叶变换及零均值标准化处理。

4.2 非独立同分布数据划分及参数设置

利用狄利克雷分布来实现 K 个客户端间的非独立同分布数据划分^[22]。具体地，首先随机采样 $p_c \sim \text{Dir}_K(\gamma)$ ，然后将 $p_{c,k}$ 比例的第 c 类样本分配至第 k 个客户端，其中 $\text{Dir}(\gamma)$ 表示浓度参数为 γ 的狄利克雷分布。两个案例中的浓度参数均设置为 0.1，值得注意的是浓度参数越小，则客户端缺失某类样本的概率越大，因此，两个案例的协作训练任务极具挑战性。图 6 展示了两个案例中各客户端的数据分布情况，其中横坐标为类标签，纵坐标为客户端编号。

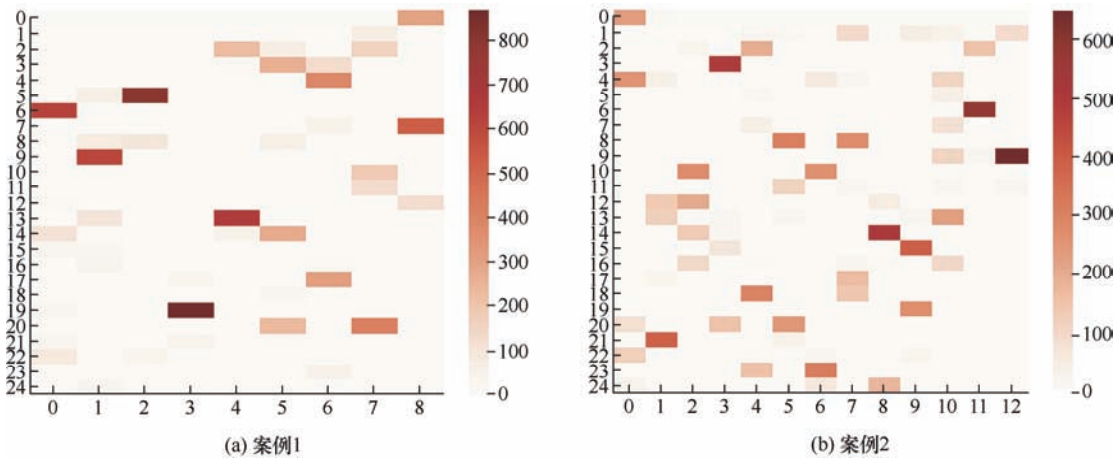


图 7 两个案例中的各客户端的数据分布

通过实验逐一确定了所提框架的超参数：全局训练总轮数 $T=30$ ，各客户端的局部训练轮数为 4，优化器为 Adam，学习率初始值为 0.001，并在全局

训练轮数达 15 时衰减为 0.000 1，权值衰减为 1×10^{-5} 。对于特征对比损失函数中的正则化常数 μ 及边缘常数 ε ，案例 1 中 $\mu=2, \varepsilon=1$ ；案例 2 中 $\mu=5, \varepsilon=0.1$ 。

对于可信任权重数 f , 假设框架内的恶意客户端数量不超过总数的 $1/3$, 因此 $f=16$ 。所有客户端的初始信誉 $\theta_0^k = 2$, 信誉最大值 $\theta^{\max} = 28$, 初始社交阈值 $h_0 = 1.5$, 其每轮增量 $\beta = 0.5$, 比例系数 $\alpha = 2$ 。

4.3 非独立同分布条件下的诊断性能测试

为测试所提框架中的特征对比损失函数解决非独立同分布问题的能力, 实验将在可信环境下进行, 并关闭框架中的拜占庭容错的评分机制以及基于信誉的激励算法。此外, 使用三种先进的联邦学习方法替换框架中的特征对比损失函数以构造三种对比方法, 包括① FedAvg^[3], ② FedProx^[10], ③ MOON^[17]。为体现协作训练的必要性, 随机令一个客户端不加入所提框架而仅使用本地数据集训练一个诊断模型, 此基线方法命名为 SOLO。每种方法进行十次重复试验, 如表 3 和图 8 所示, 采用平均诊断准确率和混淆矩阵来展示实验结果, 图 8 中横坐标为预测健康状态, 纵坐标为实际健康状态。

表 3 两个案例中各方法的平均准确率(%)

方法	案例 1	案例 2
特征对比损失函数	96.04±1.31	90.53±1.31
FedAvg	93.22±3.20	87.74±1.27
FedProx	94.62±1.32	87.91±1.62
MOON	92.81±3.51	88.67±1.26
SOLO	70.39±2.24	42.07±2.44

如表 3 所示, SOLO 的准确率远远低于其他联邦学习方法, 这验证了联邦学习的必要性。通过比较不同的联邦学习方法可以发现, 所提方法在两个案例中都达到了最高的诊断准确率, 特别是与 FedAvg 相比, 所提方法仅引入了一个额外的对比损失项, 但实现了 2.82% 和 2.79% 的准确率提升。此外, 图 8 表明所提方法在两个案例中对行星齿轮箱的各种健康状态均提供了较高的诊断准确率。

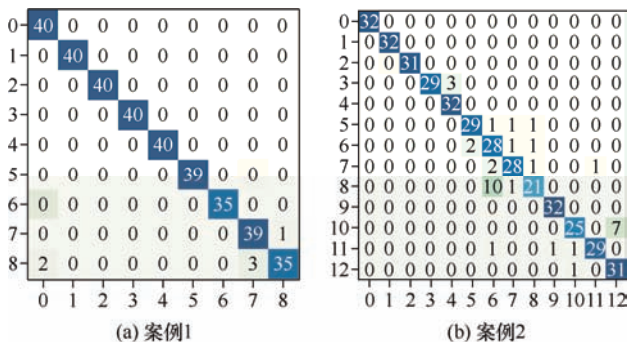


图 8 所提方法在两个案例中的混淆矩阵图

4.4 对投毒攻击的抵抗力评估

为评估所提框架对投毒攻击的抵抗力, 需随机

令一些客户端使用标签翻转生成一定数量的毒性样本, 即将样本的正确源标签修改为错误的目标标签, 且这些客户端会使用此毒性样本进行本地训练。根据标签翻转的类型, 恶意客户端的数量, 以及所制造的毒性样本占框架内该类样本总数的比例, 共设置了六项诊断任务, 其具体细节列于表 4。

表 4 投毒攻击防御任务设置

任务	标签翻转		恶意客户端的数量	毒性样本的比例(%)
	源标签	目标标签		
1	0	2	2	30
2	0	2	4	40
3	0	2	8	50
4	4	7	2	30
5	4	7	4	40
6	4	7	8	50

为凸显拜占庭容错的评分机制的有效性, 框架中的基于信誉的激励算法将被关闭。此外, 使用 Multi-KRUM 算法^[8]替换框架中的此机制来建立对比方法, 基线方法通过移除框架中的此机制构造, 并将其命名为“不抵抗”。每个方法在每项诊断任务中都进行十次重复试验, 图 9 和图 10 分别展示了两个案例中三种方法在六项诊断任务中的平均准确率。表 3 给出的所提方法在可信环境下的诊断准确率可作为比较基准, 分别为 96.04% 和 90.53%。

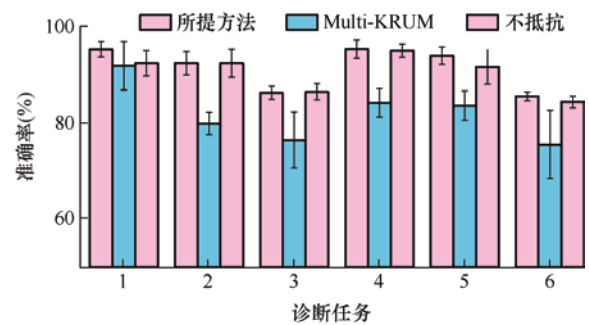


图 9 案例 1 各方法在六项诊断任务中的平均准确率

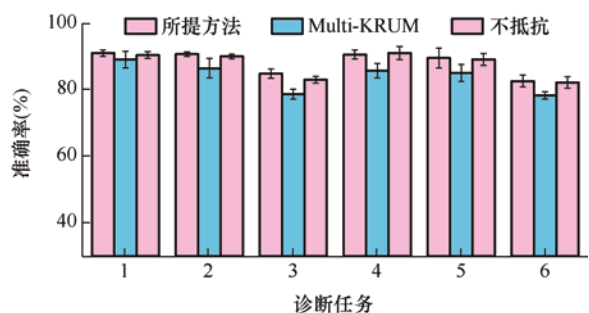


图 10 案例 2 各方法在六项诊断任务中的平均准确率

如图 9 和图 10 所示, 随着恶意客户端数量和毒

性样本比例的增加, 基线方法的诊断准确率呈现显著的下降趋势。特别是, Multi-KRUM 算法的准确率甚至会低于基线方法, 其原因是在非独立同分布条件下, Multi-KRUM 算法可能会将远离全局权重的良性局部权重误判为恶意权重并直接将其移除, 从而丢失有效信息。然而, 所提方法通过评估局部权重之间的交互性来为每个权重设置一个置信度, 尽可能筛选有效信息并减少不正确信息的影响。即使中毒样本比例高达 40%, 所提方法的准确度仍接近比较基准。此外, 在实验中观察到所提方法有可能为恶意权重分配较高的得分, 这可能是因为该权重所包含的有效信息比错误信息更多。

4.5 激励算法的合理性验证

通过将所提的基于信誉的激励算法与现有文献中的两种方法作对比来体现其合理性。这两种方法分别从本地训练的时间消耗^[15]和本地数据集的大小^[16]的角度来计算应奖励给客户端的代币数量。图 11 展示了良性客户端和恶意客户端在所提方法的训练过程中的信誉变化, 同时, 图 12 提供了这些客户端在三种方法中可获取的代币数量信息。

如图 11 所示, 对于在整个训练过程中保持良好训练行为的客户端 C^1 和 C^2 , 他们的信誉会持续增加直到达到最大值。对于客户端 C^3 , 他在训练初期发起了投毒攻击, 因此被平台隔离了数轮, 在其声誉回升到社交阈值后, 他开始使用正确样本进行训练, 因此他的声誉最终也在社交阈值之上, 并可获得一定的代币奖励。然而, 对于客户端 C^4 , 他在训练初期发起了投毒攻击并在隔离阶段结束后继续投毒, 因此该客户端几乎一直处于被系统隔离的状态。

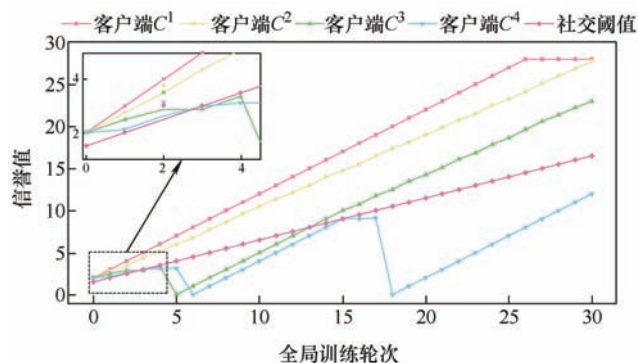


图 11 各客户端的信誉变化

如图 12 所示, 所提方法通过评估客户端在训练过程中的表现, 客观公平地为客户端发放代币奖励。特别是, 对于始终保持恶意的客户端 C^4 , 他将不会从系统中获得任何奖励。然而, 由于对比方法将本地训练的时间消耗和本地数据集的大小作为奖励大

小的衡量标准, 客户端 C^4 反而可能获得高额代币奖励, 这显然是不合理的。

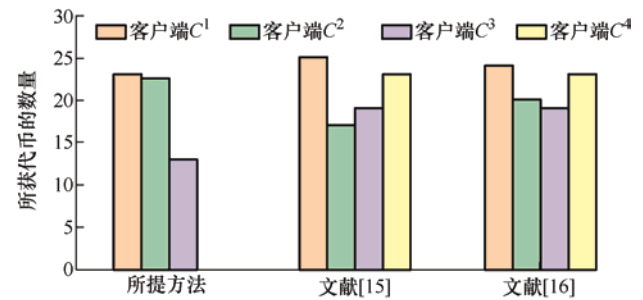


图 12 各客户端在不同评价角度中所获代币数

5 结论

提出了一种区块链和边缘计算赋能的联邦学习故障诊断框架, 在保障工业物联网中各节点的本地数据隐私和安全的前提下, 实现各节点协同训练故障诊断模型。

(1) 利用区块链技术赋予了联邦学习去中心化, 防篡改, 可审计等多种优良特性。特别是, 数据传输和模型的聚合操作是由矿工及被选中的领导者执行的, 从而避免了单点故障问题。

(2) 构造了一种特征对比损失函数, 通过在局部训练阶段限制局部模型与全局模型之间的偏移, 减轻非独立同分布条件对联邦学习性能的影响。

(3) 设计了一种拜占庭容错的评分机制, 根据各局部权重与其他局部权重的交互性来为每个权重设置一个置信度, 在非独立同分布条件下削弱恶意节点对联邦学习模型的负面作用。

(4) 开发了一种基于信誉的激励算法, 采用社交准则约束各节点的行为并评估应给予他们的代币奖励, 有效激励良性节点贡献其资源并阻断恶意节点加入训练。

参 考 文 献

- [1] 雷亚国, 贾峰, 孔德同, 等. 大数据下机械智能故障诊断的机遇与挑战[J]. 机械工程学报, 2018, 54(5): 94-104.
LEI Yaguo, JIA Feng, KONG Detong, et al. Opportunities and challenges of machinery intelligent fault diagnosis in big data era[J]. Journal of Mechanical Engineering, 2018, 54(5): 94-104.
- [2] 邵海东, 肖一鸣, 颜深, 等. 仿真数据驱动的改进无监督域适应轴承故障诊断[J]. 机械工程学报, 2023, 59(3): 76-85.

- SHAO Haidong, XIAO Yiming, YAN Shen, et al. Simulation data-driven enhanced unsupervised domain adaptation for bearing fault diagnosis [J]. *Journal of Mechanical Engineering*, 2023, 59(3): 76-85.
- [3] ZHANG W, LI X. Federated learning for machinery fault diagnosis with dynamic validation and self-supervision[J]. *Knowledge-Based Systems*, 2021, 213: 106679.
- [4] MCMAHAN H B, MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data[C]// *Proceedings of the 20th International Conference on Machine Learning*, (PMLR), 2017, 54: 1273-1282.
- [5] 方晨, 郭渊博, 王一丰, 等. 基于区块链和联邦学习的边缘计算隐私保护方法[J]. *通信学报*, 2021, 42(11): 28-40. FANG Chen, GUO Yuanbo, WANG Yifeng, et al. Edge computing privacy protection method based on blockchain and federated learning[J]. *Journal on Communications*, 2021, 42(11): 28-40.
- [6] 邱天晨, 郑小盈, 祝永新, 等. 面向非独立同分布数据的联邦学习架构[J]. *计算机工程*, 2023, 49(7): 110-117. QIU Tianchen, ZHENG Xiaoying, ZHU Yongxin, et al. Federated learning architecture for non-IID data[J]. *Computer Engineering*, 2023, 49(7): 110-117.
- [7] MA Z, MA J, MIAO Y, et al. ShieldFL: Mitigating model poisoning attacks in privacy-preserving federated learning[J]. *IEEE Transactions Information Forensics Security*, 2022, 17: 1639-1654.
- [8] ZHAO Y, ZHAO J, JIANG L, et al. Privacy-preserving blockchain-based federated learning for IoT devices[J]. *IEEE Internet of Things Journal*, 2021, 8(3): 1817-1829.
- [9] ZHAO Y, LI M, LAI L, et al. Federated learning with non-IID data[EB/OL]. arXiv preprint arXiv: 1806.00582, 2018.
- [10] KARIMIREDDY S P, KALE S, MOHRI M, et al. Suresh. Scaffold: Stochastic controlled averaging for on-device federated learning[C]// *Proceedings of the 37th International Conference on Machine Learning*. (PMLR), 2020.
- [11] ZHANG W, LU Q, YU Q, et al. Blockchain-based federated learning for device failure detection in industrial IoT[J]. *IEEE Internet of Things Journal*, 2021, 8(7): 5926-5937.
- [12] LI Y, CHEN C, LIU N, et al. A blockchain-based decentralized federated learning framework with committee consensus[J]. *IEEE Network*, 2021 35(1): 234-241.
- [13] XU Y, LU Z, GAI K, et al. BESIFL: Blockchain empowered secure and incentive federated learning paradigm in IoT[J]. *IEEE Internet Things Journal*, 2023, 10(8): 6561-6573.
- [14] LIU Y, PENG J, KANG J, et al. A secure federated learning framework for 5G networks[J]. *IEEE Wireless Communications*, 2020, 27(4): 24-31.
- [15] KANG J, XIONG Z, NIYATO D, et al. Incentive mechanism for reliable federated learning: A joint optimization approach to combining reputation and contract theory[J]. *IEEE Internet of Things Journal*, 2019, 6(6): 10700-10714.
- [16] FENG S, NIYATO D, WANG P, et al. Joint service pricing and cooperative relay communication for federated learning [C]// *iThings/GreenCom/CPSCoM/SmartData*, 2019: 815-820.
- [17] LI Q, HE B, SONG D. Model-Contrastive federated learning[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021: 10713-10722.
- [18] ZHANG Y, SCHAAR M. Reputation-based incentive protocols in crowdsourcing applications[C]// *2012 Proceedings IEEE INFOCOM*, 2012: 2140-2148.
- [19] GILAD Y, HEMO R, MICALI S, et al. Algorand: Scaling Byzantine agreements for cryptocurrencies[C]// *Proc ACM Symp Oper Syst Principles (SOSP)*, 2017: 51-68.
- [20] 陈晓磊, 徐小力, 吴国新. 物联网架构下风力发电机组远程状态监测系统[J]. *风机技术*, 2013(1): 63-66, 88. CHEN Xiaolei, XU Xiaoli, WU Guoxin. Design of Internet of Thing structured remote condition monitoring system for wind turbines[J]. *Chinese Journal of Turbomachinery Chinese Journal of Turbomachinery*, 2013(1): 63-66, 88.
- [21] LI T, ZHOU Z, LI S, et al. The emerging graph neural networks for intelligent fault diagnostics and prognostics: A guideline and a benchmark study[J]. *Mechanical Systems and Signal Processing*, 2022, 168: 108653.
- [22] YUROCHKIN M, AGARWAL M, GHOSH S, et al. Bayesian nonparametric federated learning of neural networks[C]// *Proceedings of the 36th International Conference on Machine Learning*, (PMLR), 2019, 97: 7252-7261.

作者简介: 邵海东(通信作者), 男, 1990 年出生, 博士, 副教授, 博士研究生导师。主要研究方向为故障诊断与寿命预测, 数据挖掘与信息融合, 工业大数据分析。

E-mail: hdsiao@hnu.edu.cn

肖一鸣, 男, 1999 年出生, 博士研究生。主要研究方向为联邦学习故障诊断, 不确定性分析。

E-mail: xiaoyim@hnu.edu.cn