面向可信联邦学习公平性的研究综述

陈颢瑜[1,2],李浥东[1,2],张洪磊[1,2],陈乃月[1,2]
(1. 北京交通大学计算机与信息技术学院,北京 100044;2. 交通大数据与人工智能教育部重点实验室,北京 100044)

摘 要: 联邦学习能够促进多方参与者之间的数据共享和协同计算,其已经成为一种流行的分布式机器学习范式. 联邦学习目前的研究主要集中在性能提升和隐私保护方面. 近年来,随着可信人工智能研究的深入,可信联邦学习的研究也受到越来越多的关注. 其中,保证联邦学习的公平性是面临的关键问题之一. 提升联邦学习的公平性能够保证客户端参与的积极性和联邦学习训练的可持续性. 然而,由于联邦学习中通常存在着数据异构性和设备异构性,传统的联邦学习方法会导致客户端之间具有很大的差异,无法保证所有参与者之间的公平,这会极大地影响用户参与联邦学习的动力. 基于此,对近年来联邦学习公平性的研究方法进行全面归纳梳理与深度探讨分析. 首先对当前联邦学习公平性研究的主要方向进行划分,并对每个方向的公平性定义与评价标准进行了解释及对比. 随后详细探讨了联邦学习公平性不同方向面临的挑战和主要解决方案. 最后对联邦学习公平性研究中常用的数据集、实验场景设置和公平评价指标进行了归纳梳理,并对未来研究方向与发展趋势进行探讨和展望.

关键词: 可信赖;联邦学习;公平性;数据异构;协同计算
基金项目: 国家自然科学基金(No.U1934220,No.U2268203)
中图分类号: TP309;TP399 **文献标识码:** A **文章编号:** 0372-2112(2023)10-2985-26
电子学报 URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20230139

Fairness in Trustworthy Federated Learning: A Survey

CHEN Hao-yu[1,2], LI Yi-dong[1,2], ZHANG Hong-lei[1,2], CHEN Nai-yue[1,2]
(1. *School of Computer and Information Technology, Beijing Jiaotong University, Beijing* 100044, *China*;
2. *Key Laboratory of Big Data & Artificial Intelligence in Transportation, Ministry of Education, Beijing* 100044, *China*)

Abstract: Federated learning is a distributed machine learning paradigm that facilitates data sharing and collaborative computing among multiple participants. Currently, research on federated learning primarily focuses on performance improvement and privacy protection. With the emergence of trustworthy artificial intelligence, the research on trustworthy federated learning methods has gained more attention, and ensuring fairness in federated learning is one of the main challenges. Improving the fairness of federated learning can motivate the enthusiasm of clients and ensure the sustainability of federated learning training. However, due to the heterogeneity of data and devices in federated learning, traditional federated learning methods may lead to significant performance differences between clients, which may hinder fairness among all participants and significantly impact the motivation of users to participate in federated learning. Based on this, this paper provides a comprehensive review of the research methods of fairness in federated learning. Firstly, we categorize the main research directions of fairness in federated learning, elaborates the definition and compares the evaluation criteria of fairness in each direction. Next, we discuss the challenges and main solutions for improving fairness in federated learning in each direction. Then, we summarize the commonly used datasets, experimental scenarios, and fairness evaluation metrics in the study of fairness. Finally, we prospectively explore the future research directions and development trends of fairness in federated learning.

Key words: trustworthy; federated learning; fairness; data heterogeneity; collaborative computing
Foundation Item(s): National Natural Science Foundation of China (No.U1934220, No.U2268203)

收稿日期:2023-02-17;修回日期:2023-08-31;责任编辑:宋甲英

1 引言

随着大数据和人工智能技术的持续发展,数据已经成为推动社会智能化发展的基本生产要素. 传统的机器学习方法需要将所有的数据集中处理,但在实际应用中,将用户数据收集到中心服务器进行训练会带来额外的安全风险. 由于数据中蕴含有大量的隐私信息,这种方式使所有参与方的数据面临着严重的隐私泄露风险. 因此,数据拥有者并不愿意将自身的原始数据进行共享,造成"数据孤岛"的现象日益严重. 为了使参与方之间在数据不共享的情况下,依然能够高效地进行协同计算,联邦学习的理论概念应运而生[1].

联邦学习的核心思想是:数据拥有者之间在不共享原始私有数据的前提下实现协同计算[2]. 具体来说,联邦学习能够使参与客户端利用自身的数据进行本地模型训练,然后仅向中心服务器上传模型的参数或梯度更新信息,避免将原始数据的直接传输. 随后,中心服务器通过将所有本地模型更新进行聚合,维护一个全局的共享模型. 通过这种方式,联邦学习不仅能实现数据共享和模型训练,同时也能够降低隐私泄露的风险. 然而,现有的联邦学习方法通常忽视了用户之间的公平性问题,这一问题极大程度上制约了联邦学习在实际应用中的发展. 随着联邦学习研究的深入,可信联邦学习(trustworthy federated learning)新范式应运而生. 可信联邦学习是一种增强型的联邦学习,在保证数据隐私和模型可证安全的同时,提高了决策机制的公平性和可解释性. 特别是在涉及隐私敏感数据的场景中,可信联邦学习能够有效降低数据泄露风险,并提供可溯源和审计监管的能力. 可信联邦学习包含了鲁棒性、公平性、隐私性、可解释性、可审计可监管和环境友好性等关键维度,其中公平性是其中最关键的维度之一.

公平是指在处理事情的过程中不偏袒任何一方,这一理念广泛体现于各种领域,并在人类的社会活动中占有重要地位[3~7]. 公平性的研究一直受到不同领域学者的广泛关注. 美国心理学家亚当斯1965年提出的公平理论指出:人的工作积极性不仅与个人实际报酬多少有关,而且与人们对报酬的分配是否感到公平更为密切. 人们总会自觉或不自觉地将自己付出的劳动代价及其所得到的报酬与他人进行比较,从而判断是否获得公平对待. 因此,保证公平性对个人的生存和社会的发展稳定具有重要意义. 公平性在机器学习中也得到了大量的研究[8~12]. 机器学习算法需要以数据驱动的方式进行学习,很可能导致在训练的过程无意中引入人类的偏见,机器学习中公平性研究的目标就是使算法消除因个人或群体属性引起的偏见[13].

作为一种多参与方的分布式机器学习范式,联邦学习公平性的研究不仅涉及机器学习算法本身,还包括服务器与客户端之间的通信传输、模型的聚合以及参与方之间的博弈等一系列过程. 这些过程都会对联邦学习的公平性造成影响,使联邦学习公平性的研究更具挑战性. 提升联邦学习的公平性能够确保联邦学习训练的可持续性和客户端参与的积极性. 传统联邦学习方法中,针对损失函数的建模是造成联邦学习不公平现象的主要原因之一. 例如联邦平均(Federated Averaging,FedAvg)算法[14]按照客户端样本比例对参与客户端的局部损失进行加权平均,以拟合大多数客户端的本地数据. 然而,联邦学习参与客户端通常存在着数据异构性和设备异构性,部分参与者可能会获得性能更好的模型,而部分参与者可能生成的模型性能较差,进而影响全局模型的更新. 另外,在用户参与联邦学习的过程中,不可避免会存在设备计算和通信等资源的消耗. 若用户无法得到与自身资源消耗相匹配的公平回报,最终可能不愿参与到联邦学习的过程中. 因此,如何提高联邦学习的公平性是一个具有挑战性的课题.

近年来,联邦学习的研究引起了广泛的关注[15],同时在公平性方面也取得了一定的进展[16~18]. 由于应用场景的不同,联邦学习公平性的需求和定义也不尽相同,目前国内缺乏对联邦学习公平性的梳理与总结. 基于此,本文对联邦学习公平性进行了全面综述,旨在为联邦学习公平性的研究梳理出清晰的研究脉络,助力于研究者针对公平性进一步的探索与研究. 本文的具体贡献如下:

(1)归纳梳理出联邦学习中的偏见、公平性定义和分类体系.

(2)根据不同的应用场景整理了相关的研究工作,并详细探讨了相应的联邦学习公平性解决方案.

(3)归纳总结了联邦学习公平性研究中常用数据集、实验场景设置和公平评价指标.

(4)指出联邦学习公平性研究面临的挑战,并展望了未来的研究方向.

本文第1节阐述了联邦学习公平性研究的背景和重要意义. 第2节对联邦学习中的偏见、公平性的定义及分类进行介绍. 第3节详细介绍基于平均分配理论的联邦学习公平性研究存在的挑战以及具体的解决方案. 第4节详细介绍基于按劳分配理论的联邦学习公平性研究存在的挑战以及具体的解决方案. 第5节归纳梳理了联邦学习公平性研究中的常用数据集实验场景设置和公平性评价指标. 第6节对未来的研究进行展望. 第7节对全文进行总结.

2　联邦学习公平性定义

本节首先介绍联邦学习的基本框架以及当前联邦学习中存在的偏见,然后在此基础上介绍了联邦学习中的公平性定义,最后对联邦学习公平性的主要研究方向进行分类探讨.

2.1　联邦学习

联邦学习是一种新兴的人工智能基础技术,其核心思想是数据拥有者在不共享原始私有数据的情况下在本地训练模型,然后将参数或梯度信息传输至服务器端,服务器端聚合本地更新并维护全局共享的模型进行协同计算. 联邦学习通过本地独立训练和全局集中聚合的分布式协同计算方法,解决了单一客户端数据量少、数据质量低导致的本地模型性能差的问题,同时实现了数据隔离和隐私保护.

联邦学习算法的本质是求解优化问题. 通用的联邦学习框架由多个客户端 $k=\{1,2,\cdots,N\}$ 组成,分别拥有由 M_i 个样本组成的数据集 D_i,每个客户端 $k\in N$ 利用其自身的数据集进行本地模型训练,优化得到局部模型参数 w 并将其上传至中央服务器,之后通过中央服务器的聚合过程得到全局模型 G. 联邦学习的目标函数是使下面的标准优化模型最小化,即

$$\min_{w\in W} f(w):=\sum_{k=1}^{N} p_k F_k(w) \tag{1}$$

对于机器学习问题,我们可以选取

$$F_k(w)=\frac{1}{\left|D_i\right|}\sum_{x\in D_i} l_k(w,x,y) \tag{2}$$

其中,N 表示参与联邦学习的客户端总数量,p_k 表示客户端 k 的聚合权重,其中 $\sum_{k}^{N} p_k=1$,$l_k(w,x)$ 表示第 k 个客户端中在给定模型参数 w 上对样本 (x,y) 进行预测所得到的损失.

典型的联邦学习模型通常包含以下步骤:

步骤 1　服务器选取客户端参与本轮的模型训练更新过程.

步骤 2　客户端从服务器端下载最新的全局更新模型参数.

步骤 3　客户端根据下载的模型参数使用自身数据集训练本地模型.

步骤 4　经过几个时间段的本地训练后,客户端将训练好的本地模型参数信息上传至服务器端.

步骤 5　服务器端对收集到的本地参数信息进行全局聚合,生成最新的全局模型.

联邦学习会重复执行步骤 1 至步骤 5,直到训练出期望的模型. 因此,本地模型训练和全局聚合规则在联邦学习过程中同样起着至关重要的作用. FedAvg[14]是目前最经典的联邦学习方法,它通过聚合来自客户端局部模型更新的平均值作为全局模型更新,其中每个客户端的聚合权重是根据其训练样本的数量比例进行加权.

然而,传统的联邦学习方法(如 FedAvg)并没有考虑公平性的问题,不能保证联邦学习中的公平性. 在实际场景中客户端之间的数据通常是非独立同分布的,本地模型的更新方向存在着很大的差异,导致客户端优化得到的本地模型更新会偏离服务器聚合得到的全局模型更新,最终模型之间的差异很大,每个客户端无法得到公平的对待,这会严重影响用户参与联邦学习的积极性. 另外,可信联邦学习也为公平性的研究带来了更深层的挑战. 在可信联邦学习中不同维度之间并不是彼此独立,而是相互关联的. 一个维度的提高可以促进另一个维度,例如,可解释性的研究能够帮助联邦学习服务器探索应该如何做出公平的决策以及这些决策如何影响每个用户的利益. 通过这种方式增加用户对联邦学习系统决策的信任和理解,从而激励用户加入联邦学习过程中. 另外,在追求公平性的过程中,可审计可监管性可以提供对算法决策的解释和验证,以确保公平性原则得到遵守. 同时,可审计可监管性可以帮助发现和纠正潜在的偏见或歧视,以确保算法的决策过程是可靠和可控的. 但同时不同维度之间也会存在着冲突,这导致在某些场景下无法同时满足两个或更多维度,甚至可能会相互损害. 例如,公平性的研究通常需要增加算法和模型的复杂性和计算成本,而这可能会导致计算资源的需求增加. 如果计算资源有限,那么在追求公平性的同时可能无法承担高计算成本,这可能导致公平性的牺牲. 另外,实现公平性可能需要对模型进行调整或增加约束限制,以减少对某些群体的错误分类或差异. 这可能会导致模型的鲁棒性下降,即对抗各种干扰和攻击的能力减弱. 因此,对可信联邦学习公平性的研究至关重要. 下一节将详细介绍联邦学习中存在的偏见、公平性的定义与内涵,期望能够为研究者们提供全面的框架和新的研究思路.

2.2　联邦学习中的偏见*

随着联邦学习研究的日益深入,保证联邦学习客户端之间的公平性已经成为一个主要的挑战. 联邦学

---

* 偏见的英文为 bias. Bias 可以翻译为偏见或偏差. 偏见表示主观造成的不公平,偏差适用于描述客观现象形成的差异. 联邦学习中同时存在着主观和客观形成的不公平. 偏差多存在客户端固有属性的差异,而偏见多存在于通信传输和全局聚合阶段,这是联邦学习给公平性带来的独特挑战. 为了体现联邦学习给公平性带来的独特挑战,同时为了行文的一致性流畅性,我们在文中统一使用了偏见.

习作为多参与方的分布式机器学习范式,不仅包括机器学习的算法,也包含了服务器与客户端之间的通信传输、模型的聚合以及参与方之间的博弈等一系列过程,这些过程都会对联邦学习的公平性造成影响. 偏见必然会导致不公平. 对联邦学习中的偏见进行梳理能够更好地去解决这些问题. 本节根据联邦学习的全过程,从数据、算法、交互和聚合等方面对联邦学习中常见的不公平现象和各种形式的偏见进行分类介绍,如图1所示. 其中,数据类和算法类的偏见与机器学习中存在着一定程度的相同,通常是客户端之间固有属性的差异客观形成的不公平现象. 交互偏见和聚合偏见是联邦学习过程所带来的公平性挑战,这类偏见通常是通信或聚合规则不合理而主观造成的用户之间不公平对待.

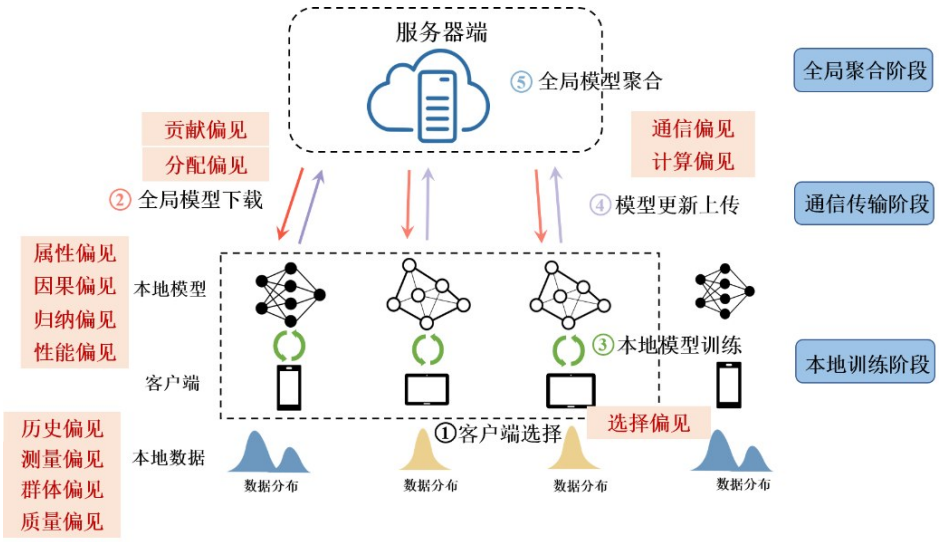


图1 联邦学习偏见类型

2.2.1 数据偏见

历史偏见. 历史偏见指的是现实世界长期自身存在的差异,例如社会、文化和习俗等方面的差异造成数据中产生的偏见. 例如,在手写字符识别中,不同作者的字迹并不相同,从而造成数据集之间天然的特征分布不平衡.

测量偏见. 测量偏见指的是每个参与者收集的数据与现实世界中真实数据之间存在差异. 在选择、收集或计算用于预测的特征和标签时,可用或能测量的数据往往是感兴趣的特征和标签的有噪声代理,在选择了要测量的代理后,测量过程本身又增加了噪声. 例如,逮捕率通常被用来代替犯罪率,然而少数族裔社区会比白人社区受到政府更多的监管,可获得测量的逮捕率作为代理导致了测量偏见.

群体偏见. 群体偏见指的是数据所表示群体统计特征和属性特征与应用目标群体不同所产生的偏见. 例如,不同软件平台上群体性别的代表性存在差异,女性在购物和聊天等软件中较为活跃,而男性在体育和游戏软件中较为活跃.

质量偏见. 质量偏见通常是由于联邦学习中每个客户端会自主生成和收集本地数据,这使客户端的本地数据集之间通常是非独立同分布的,从而造成客户端之间的数据质量通常存在着偏见,这种差异会导致每个客户端的本地模型和实际贡献也并不相同.

2.2.2 算法偏见

属性偏见. 属性偏见通常发生在选择和利用属性的过程中,指某些敏感属性(如性别、种族)形成的不同群体之间算法决策的差异. 对属性的排除、包含和加权等操作均可能引起机器学习算法的偏见. 面向不同任务,相同的属性变量应采取不同的处理方式以适应任务.

因果偏见. 因果偏见通常是将关联关系误认为因果关系所导致的偏见,造成因果关系构建不合理. 例如,智力会影响受教育和收入的程度,在构建机器学习模型时引起因果偏见,即认为受过更多教育的人赚更多的钱仅仅是因为他们更聪明,而不是因为他们受过更多的教育;在学校接受辅导的同学的考试成绩比没有接受辅导的同学的考试成绩差,接受辅导并不是考试成绩差的原因,考试成绩差是需要接受辅导的原因.

归纳偏见. 归纳偏见发生在机器学习算法的测试评估阶段. 机器学习算法的目标函数通常设定为最小化均方误差,那么如果从样本数量的角度理解,拟合多数群体比拟合少数群体更重要(对极小化误差更有利). 极端情况下,与多数群体的数据分布显著不同的少数群体可能被视为离群数据样本.

性能偏见. 性能偏见指的是客户端通过联邦学习获得模型最终性能之间的差异. 算法不合理造成每个参与客户端之间模型性能存在偏见,算法以牺牲其他模型为代价过度拟合任何特定模型. 与机器学习不同,联邦学习中包含了多方参与者共同协作,客户端之间模型性能的公平至关重要.

2.2.3　交互偏见

通信偏见. 通信偏见是指联邦学习中,客户端的通信传输能力不同,导致不同客户端与服务器之间的通信存在差异. 例如,在存在大规模客户端的场景中,服务器端在全局聚合过程中会接入大量客户端,同时接收太多设备的反馈会导致服务器端网络拥塞的问题. 此外,由于客户端的通信能力和电池时间有限,任务调度会因设备而异,因此很难在每个更新轮次结束时精准地同步接入所有客户端.

计算偏见. 每个客户端的计算能力不同,这会导致每个客户端本地模型训练时间的不同. 在每轮的训练过程中需要等待最慢的客户端训练结束才能进行上传聚合,这会严重影响联邦学习的效率,对训练快的客户端也是不公平. 若规定时刻上传,服务器无论是直接舍弃超时客户端的本地更新,还是将客户端未完成的本地更新直接上传,都会对最终全局聚合模型的更新造成很大影响.

2.2.4　聚合偏见

选择偏见. 选择偏见指的是每个客户端在聚合过程中被选择几率之间的偏见. 服务器端在每轮的全局聚合过程中会选择客户端的参数更新信息进行聚合,每个客户端被选择的机会不相同造成差异. 为了实现客户选择的公平性,客户端需要获得公平的选择机会,在服务器端的利益和客户端的利益之间取得平衡.

贡献偏见. 贡献偏见指的是贡献评估值与客户端的实际贡献值之间的偏见. 联邦学习中每个参与客户端的贡献程度并不相同,需要对每个客户端的实际贡献进行精准评估. 直观上看,拥有数量大和质量好数据集的客户端应该能够做出更高的贡献,贡献更多的客户端应该获得更好的回报.

分配偏见. 分配偏见指的是参与客户端收获的奖励无法与其实际贡献相匹配. 传统联邦学习中服务器端通过联邦平均等方法聚合得到的全局模型参数可能会在不同客户端之间造成偏见,显然每个客户端获得相同的模型参数对提供数据资源较多的客户端是不公平的.

偏见不可避免地会导致不公平. 只有明确了联邦学习中存在的各种偏见,才能更好地设计方法消除这些偏见,从而提升联邦学习的公平性. 虽然联邦学习中的偏见可以粗略地分为上述四大类,但这些分类可能并没有完全涵盖联邦学习中可能出现的所有偏见. 在下文中,我们将针对上述偏见探讨目前具体提高公平性的解决方案. 值得注意的是,在当前的联邦学习公平性研究中,某些数据偏见和算法偏见与机器学习的公平性问题存在相似之处. 因此,我们将重点探讨联邦学习过程中给公平性带来的特殊挑战.

2.3　联邦学习的公平性

随着联邦学习研究的日益深入,保证联邦学习客户端之间的公平性已经成为一个主要的挑战. 在实际场景中,联邦学习具有不同的公平性描述和定义. 根据研究目标的不同,我们将现有的联邦学习公平性研究划分如下:

(1) 公平性目标是减少客户端之间的分配差异,使每个参与客户端最终获得的收益更加均衡(我们称之为基于平均分配的联邦公平学习).

(2) 公平性目标使每个参与客户端获得的计算回报与其在联邦学习过程中的实际贡献相匹配(我们称之为基于按劳分配的联邦公平学习).

平均分配的理论旨在使每个参与客户端最终获得相同的收益. 传统的联邦学习并没有考虑公平性的问题,聚合后的模型可能会偏向某些属性或客户端,特别是当客户端的训练数据自身就存在偏见时. 这种情况可能会使性能较差的客户端模型越来越差,造成客户端模型之间性能的不平衡. 基于平均分配理论的联邦公平学习研究目标是减少客户端之间的分配差异,同时保持收益的均衡,通过算法优化等方法以更小的代价保证联邦学习的公平性.

平均分配理论中公平性的评判标准为[19]: 如果w_1在客户端之间的测试性能比w_2更均匀,则表示模型w_1比w_2更公平,即

$$\mathrm{std}\left\{F_k\left(w_1\right)\right\} < \mathrm{std}\left\{F_k\left(w_2\right)\right\} \tag{3}$$

其中,$F_k(\cdot)$表示客户端k中的测试损失,$\mathrm{std}(\cdot)$表示客户端之间的标准差.

按劳分配的理论旨在使参与客户端获得的计算回报(如货币奖励、资源奖励或模型性能等)与其在联邦学习中的实际贡献相匹配. 在实际的数据异构场景中,部分参与者可能促进全局优化过程,而部分参与者可能会损害全局优化过程,导致模型的贡献和性能在不同客户端之间会有很大差异. 这种情况下若还是以平均分配理论的思想使每个客户端获得相同的收益,则对在训练过程中贡献更多的客户端是不公平的,这样会极大地损害他们参与联邦学习的动力. 为此,基于按劳分配的联邦公平学习研究目标是参与客户端最终会获得与其在联邦学习中的实际贡献相匹配的回报.

按劳分配理论中公平性的评判标准为：在联邦学习中贡献高的参与者应该获得比贡献低的参与者更好的回报，在数学表达上目前尚未形成统一的标准. Lyu等人[20]将其定义为：公平程度可以通过参与者的实际贡献与获得回报之间的相关系数来量化，并将客户端参与联邦学习的最终模型性能当作贡献回报. 客户端的本地模型性能越好通常表示在联邦学习中具有更多的贡献. 具体地，x_k表示客户端k若不参加联邦学习时本地的模型性能，最终汇集成客户端性能的集合x，即

$$x=\{x_1,x_2,\cdots,x_N\} \tag{4}$$

同样地，y_k表示客户端k最终的模型性能，最终汇集成客户端性能的集合y，即

$$y=\{y_1,y_2,\cdots,y_N\} \tag{5}$$

因此，它们之间的相关系数（如皮尔逊相关系数）就表示联邦学习方法的公平程度，即

$$F_{xy}=\frac{\sum_{k=1}^{N}(x_k-\bar{x})(y_k-\bar{y})}{(N-1)s_x s_y} \tag{6}$$

其中，$\bar{x}$和$\bar{y}$分别表示x和y的均值，s_x和s_y表示x和y的标准差，F_{xy}的值越高代表越公平. 这种相关系数的方法同样能扩展到货币奖励和资源奖励的过程中.

平均分配理论主要是为了解决联邦学习客户端之间存在的质量偏见、性能偏见、通信偏见、计算偏见和选择偏见等. 按劳分配理论主要是为了解决联邦学习客户端之间存在的质量偏见、通信偏见、计算偏见贡献偏见、分配偏见和选择偏见等. 具体情况如图2所示.

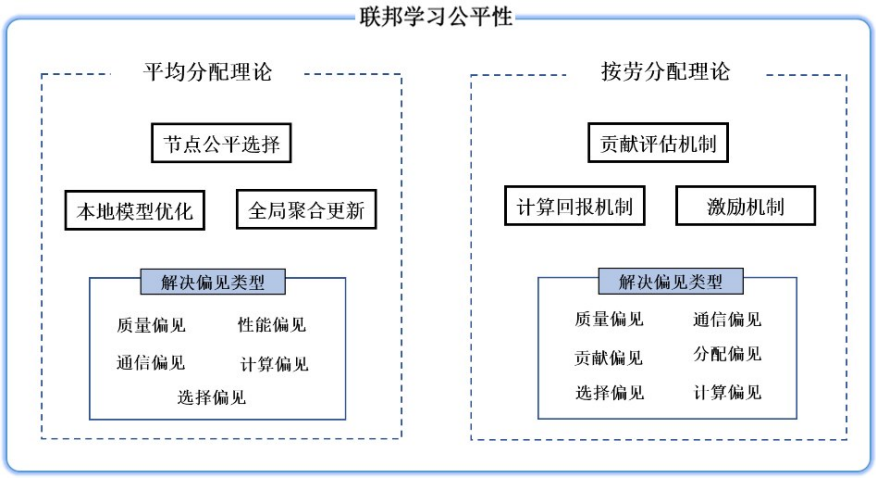


图2 联邦学习公平性分类及解决方案

表1对比分析了两类方法之间的主要思想、研究目标和实现机制. 图3呈现了联邦学习公平性技术的发展历程. 自2016年联邦学习的框架被提出以来，联邦学习的公平性问题受到了广泛的关注与研究. 经典的联邦学习框架FedAvg也可被看作一种公平性方法，尽管它的研究初衷并不是公平性问题，但其中包含了一定程度上的均衡思想. 2019年，Mohri等人[21]和Li等人[22]相继提出了基于平均分配理论的联邦学习公平性，旨在减少客户端之间的分配差异，促进联邦学习中客户端的收益更均衡. 2020年，Lyu等人[20]提出了基于按劳分配理论的联邦学习公平性理论，强调参与客户端获得的计算回报应与其在联邦学习过程中的实际贡献相匹配. 这些新探索推动着联邦学习公平性朝着方案多样化和场景完善化的方向不断发展. 基于上述的划分情况，下文将详细介绍平均分配理论和按劳分配理论的思想内涵、研究挑战和解决方案，希望为研究者们提供一个清晰明确的研究框架.

3 平均分配理论

平均分配理论的思想是每个客户端最终的收益相同，在联邦学习中客户端的收益可以通过该客户端模型的最终性能体现. 为了反映客户端之间性能的差异情况，平均分配理论的公平评价指标通常是客户端模型性能之间的标准差. 经典的联邦学习框架FedAvg通过加权平均的方法对客户端的本地模型更新进行聚

表1 联邦学习公平性研究方向汇总

公平定义	评价标准	研究目标	实现机制	代表文献
平均分配	标准差、方差、基尼系数等	客户端之间获得均衡的收益	节点公平选择、本地模型优化、全局聚合更新	文献[23~29]
按劳分配	皮尔逊相关系数、Jain公平指数等	客户端收益与实际贡献相匹配	贡献评估机制、公平回报机制、激励机制	文献[20,30~35]

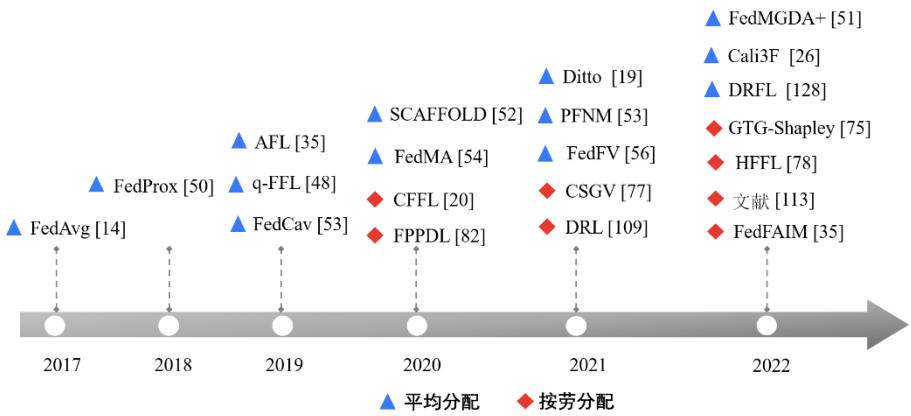


图 3　联邦学习公平性发展历程

合,也可以被看作为平均分配理论在联邦学习中的一种方法.

3.1　研究挑战

平均分配理论的研究目标就是减少客户端之间的差异,使其都能最终获得相同的收益. 为了达成这个目标至少存在如下几个挑战需要得到有效解决.

3.1.1　如何减少客户端模型最终收益之间的差异

在实际场景中客户端之间的数据通常是非独立同分布的,每个客户端本地训练后的模型可能会偏离全局模型的优化方向,造成客户端模型之间性能的不平衡. 如何通过本地模型算法或全局聚合规则的设计使模型之间能够消除性能的差异,使全局模型不会偏向某一方客户端的本地模型,最终所有客户端能够获得相同的收益,是平均分配理论研究的核心.

3.1.2　如何寻求精度、效率和公平之间的平衡

在减少模型收益差异的过程中,不管是设计模型优化算法还是修改全局聚合规则,不可避免地会对其中一些客户端的性能造成影响,从而影响联邦学习整体的精度或效率. 因此,如何在算法的精度、效率和公平性之间取得一个很好的平衡是平均分配理论研究的关键.

在本节中我们将介绍目前平均分配理论常用的解决方案,按照联邦学习的流程将其划分为:节点公平选择、本地模型优化和全局聚合更新的方法. 在节点选择阶段常用的方法有选择控制和动态补偿等;在本地训练阶段常用的方法包括在客户端本地目标函数中进行权重优化、增加正则项或增加公平性约束等;在全局聚合阶段常用的方法有修改客户端和中心服务器之间上传下载的参数更新或修改中心服务器的全局聚合更新规则. 具体的技术分类如图 4 所示.

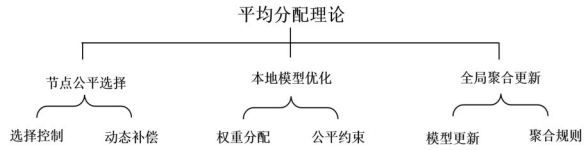


图 4　平均分配理论公平性的技术分类

3.2　节点公平选择

节点公平选择是联邦学习公平性研究中的一个关键方面,用于解决联邦学习中存在的选择偏见和通信偏见[36, 37]. 服务器端在每轮的全局聚合过程中会选择客户端的参数更新信息进行聚合,每个客户端被选择的机会决定了客户端在训练中的参与度,对确定最终模型对不同数据的偏向程度具有至关重要的作用,常用的方法包括选择控制和动态补偿.

公平的节点选择机制并不意味着以相同的概率选择每个客户端. 由于客户端之间存在的设备异构性和数据异构性,一些客户端可能需要更长的时间进行模型训练,同步模型聚合会应为训练速度慢的客户端. 为此,现有的节点选择机制通常会根据客户端的传输速度、带宽、本地精度等指标对节点的选择进行限制. 具体地,Takayuki Nishio 等人[38]提出了 FedCS 方法,根据客户端的通信能力、计算能力以及相关的数据资源大小为依据进行节点选择. Ribero 等人[39]中计算每个客户端的更新信息量,并在每轮的训练过程中选择更新信息量大的客户端进行传输. Wang 等人[40]提出了一个基于强化学习的经验驱动的联邦学习框架 FAVOR,它可以智能地选择每轮参与联邦学习的客户端设备,以抵消非独立同分布数据(Non-IID)引入的偏见. 这些方法通常会根据相应的指标在联邦学习的过程中设置客户端的准入阈值,通过这种方法过滤掉低于阈值的客户端,确保参与的客户端具有更高的代表性. 然而,这类的节点选择方法存在一些问题:一方面,有些方法只考虑整体收益的最大化,无法保证每个客户端在联邦学习中获得公平的选择机会;另一方面,有些方法旨

在加快联邦学习训练速度,服务器端会优先选择计算能力和通信能力更好的设备,而条件较差的客户端将被排除在协作训练之外,这会导致每个客户端无法获得公平的选择机会,从而使最终训练得到的模型会偏向于某些性能好的客户端模型.因此,为了实现公平的节点选择,需要综合考虑服务器端和客户端的实际情况,在服务器端利益和客户端利益之间取得平衡,以确保每个客户端能够获得更公平的选择机会.

为此,研究者提出了新的选择控制方法,强调提升不经常被选择的客户端参与训练的频率,从而减少对联邦学习中弱势客户端的选择偏见.Yang等人[41]研究了采样约束的设计,考虑了联邦学习中客户端的参与频率,并使不经常被选择的客户能够获得更大的选择优先级,从而能够更频繁地参与联邦学习的过程,促进每个客户端的选择公平.针对客户端之间的数据异构性和设备异构型,Li等人[42]提出了PyramidFL框架进行细粒度的客户端选择.该方法不仅分析了被选择和未被选择参与者之间的差异,而且利用客户端的数据和设备差异性评估了客户端的效用进一步优化选择方法.Huang等人[43]提出了一种RBCS-F的联邦学习客户端选择方法,在保证公平性的同时提高联邦学习的训练效率和准确性.该方法设置了恒定的参数作为公平性约束条件,采用排队动态方法来优化客户端的参与率,确保每个客户端的平均参与率不低于预期阈值,最终实现联邦学习中客户端的公平选择.之后,Huang等人[44]进一步改进了该方法,采用Exp3算法来计算每轮客户端的选择概率,通过这种方法实现了动态优化的公平性参数.Sultana等人[45]设计了Eiffel方法自适应地调整本地和全局模型更新的频率,在降低通信开销的同时提升了联邦学习的公平性.通过对选择控制方法的分析可以发现,采用更公平的节点选择机制可以在一定程度上提高最终模型的准确性,但是在方法设计的时候需要考虑更多的约束条件,避免训练效率的大幅降低.

设计动态补偿机制也是提高联邦学习公平性的一种策略.在该机制中,未被选中的客户端可以获得动态补偿来使其获得相应的训练.Wang等人[46]提出了一种基于端到端通信的方法PRLC,该方法采用部分设备参与全局模型更新的方式,而未参与更新的设备可以通过PRLC 在本地进行更新以减少与全局模型的偏见.节点公平选择的核心目标是在保证效用的同时,使全局公平最大化,而动态资源分配的优化对实现公平最大化至关重要.为了解决客户端之间数据不均衡造成的选择偏见,Hao等人[47]提出了零样本数据增强技术减少数据的统计异质性.另外,为了避免弱势的客户端被排除在训练过程之外,还可以使用为这些客户端提供更多权重的方法.这类方法也常被应用于本地模型优化的过程中.

3.3 本地模型优化

联邦学习的本质是一种分布式的机器学习方法,在客户端的本地模型训练阶段,可以通过对每个客户端模型算法的设计,使模型在本地训练过程中进行优化,减少模型之间的差异,从而实现每个客户端之间最终模型性能的平均分配公平.

本地模型优化主要是为了解决联邦学习中的计算偏见和性能偏见.它的主要思想是设计联邦学习模型的目标函数,使模型在训练的过程中不断优化,减少客户端之间模型的差异,从而实现模型性能的均衡,这里的性能通常指的是模型的预测精度.具体的目标函数如下所示:

$$\min_{w} \max_{k} F_k(w) \tag{7}$$

前面我们已经分析了在数据异构性和设备异构性的情况下,本地模型与全局模型之间会存在偏见,使本地模型的更新方向可能会朝着偏离全局模型的方向优化,最终导致本地模型之间的性能存在很大差异.为了减少客户端之间的差异,当前的研究通过在客户端本地训练阶段为目标函数重新分配权重或增加正则项约束并优化每个客户端模型的更新方向,这类方法是缓解联邦学习中模型之间性能不平衡的一种直接解决方案.

为了减少模型的偏见,使全局模型训练结果不会偏向于某些客户端的模型,同时使每个客户端都有机会获得较好的模型性能,Mohri等人[21]提出不可知联邦学习(Agnostic Federated Learning,AFL)框架,该方法每次会优化性能最差的客户端,避免这些客户端的性能越来越差,同时防止全局模型偏向于某些性能好的客户端模型,它的目标函数如下所示:

$$\min_{w} \max_{\lambda} \sum_{k=1}^{N} \lambda_k F_k(w) \tag{8}$$

其中,λ_k是客户端k的权重.在非独立同分布场景中,不同客户端的数据分布可能与全局目标的分布并不一致,AFL算法能够使全局模型对任何情况的客户端目标分布进行优化.只要其他客户端的优化过程不增加性能最差客户端的损失,那么他们的模型性能就不会受到这种优化算法产生的负面影响.总的来说,AFL算法的核心就是最大化联邦学习中最差客户端的性能.这种做法在一定程度上会提升整体性能,然而也会潜在搁置所有其他客户端的性能优化,使只有性能较差的部分客户端会得到性能的提升.另外,该算法缺乏灵活性,在具有小规模参与方的联邦学习中能取得不错的效果,但是难以适用于大规模的联邦学习过程中.

联邦学习的训练目标是最小化所有客户端的损失,为了使最后的所有客户端的损失最小,可能会牺牲某些客户端的性能,导致某些客户端的模型实际效果很差. 为此,Li等人[48]提出了q-公平性联邦学习(q-Fair Federated Learning,q-FFL)算法促进联邦学习中客户端的性能分布更均衡,它的目标函数如下:

$$\min_{w} F_q(w) = \sum_{k=1}^{N} \frac{p_k}{q+1} F_k^{q+1}(w) \tag{9}$$

该算法的核心是参数q的设计,这里的q起到了权重再分配的作用. 具有较高损失的设备将会被分配更高的相对权重,从而能够获得更多的优化机会以此来减少模型之间的差异. 特别地,如果参数q设置得足够大,那么q-FFL算法就等同于AFL算法,所以AFL可以当作q-FFL中的一种特例. q-FFL能够在尽可能不损失总体利益并保持最后模型准确率的前提下,将客户端之间准确率的方差减小45%,使客户端之间的性能分布更加均衡. 此外,Li等人还设计了一个可扩展的q-FedAvg方法,解决联邦学习公平性问题.

增加公平性约束的优化算法同样也是平均分配理论公平性研究的关键技术. 这类方法的思想是,通过对目标函数增加公平性约束,减少相应优化问题的可行集大小,从而排除不公平解以满足目标公平性约束. 公平性约束的联邦学习优化问题可以表示如下:

$$\begin{aligned} &\min l_{\text{utility}} \\ &\text{s.t. fairness contraint}(s) \end{aligned} \tag{10}$$

其中,约束项是指不允许出现的对应组合对应最小项. 针对AFL算法存在的不足,Du等人[49]提出了满足公平性的不可知联邦学习(Agnostic-Fair)算法应对所有情况下数据分布未知的挑战,并将其设定为学习者和对抗者之间的极大极小对抗博弈. 对抗者的目标在任何可能的未知数据分布中最大化分类器损失,学习者的目标是最小化损失. 为此,该算法对单个样本进行重新分配权重来解决不同客户端的数据分布偏移问题,随后在损失函数中增加公平性约束并使用核函数参数化算法,保证了联邦学习在未知数据分布中的公平性和最终模型的性能.

除此之外,增加正则项也是一种很好的解决方案. 正则化的联邦学习公平性优化问题可以表示如下:

$$h = l_{\text{utility}} + \lambda l_{\text{fair}} \tag{11}$$

其中,l_{utility}和l_{fair}分别是本地目标函数和公平的优化目标,λ控制着正则项l_{fair}的公平影响程度.

联邦近端(Federated Proximal,FedProx)算法[50]是一种经典的正则项优化算法,可以看作对FedAvg的泛化和重构. 相较于FedAvg,FedProx主要有两点改进:

(1)引入正则项优化客户端模型更新的方向,从而限制数据异构性造成的模型之间的差异.

(2)客户端不再需要训练相同的轮数,只需要得到一个相对的不精确解. 这种方式能够有效缓解设备的计算压力,提高联邦学习的训练效率,它的目标函数如下:

$$c_k^t(w, w^t) = l_k(w) + R_k(w, w^t) \tag{12}$$

其中,$l_k(w)$是客户端k的本地目标函数,w^t表示第t轮的全局模型,w表示第$t+1$轮本地模型求解的参数,$R_k(w, w^t)$是算法设计增加的正则项,它的具体表示为

$$R_k(w, w^t) = \frac{\mu}{2}\|w - w^t\|_2 \tag{13}$$

FedProx算法在客户端本地目标函数中增加一个公平性正则项,用以优化本地模型的更新方向,从而使每个客户端在本地训练后得到的模型w向初始时的全局模型w^t方向靠近. 通过这种方法,客户端每轮的模型更新都会朝着一个方向靠近,从而能够减少模型之间的差异. 另外,FedProx能够允许局部训练轮数的变化,不需要在训练前手动为每个客户端分配训练的轮数,从而在一定程度上解决了设备异构性的问题,提升了联邦学习的训练效率.

随着联邦学习公平性研究的进一步深入,人们对算法的鲁棒性有了更多的要求. 为此,FedMGDA+和Ditto算法相继提出,它们能够在提升联邦学习公平性的同时保证算法的鲁棒性,使算法能够抵御某些恶意设备的攻击. FedMGDA+[51]将多目标优化方法推广到联邦学习场景下,并通过修改客户端的聚合权重降低全局模型对部分客户端的偏好. 最终模型在不牺牲任何参与客户端的前提下收敛到帕累托最优,保证了联邦学习的公平性和鲁棒性.

Ditto[19]将多任务学习应用在联邦学习中,它的目标函数如下:

$$\begin{aligned} &\min_{v_k} \; h_k(v_k;\, w^*) := F_k(v_k) + \frac{\lambda}{2}\|v_k - w^*\|^2 \\ &\text{s.t. } w^* \in \operatorname*{argmin}_{w} G\left(F_1(w), F_2(w), \cdots, F_N(w)\right) \end{aligned} \tag{14}$$

其中,$F_k(\cdot)$是设备k的本地优化目标,v_k是设备k的个性化模型,$G(\cdot)$是全局聚合规则,w^*是全局模型参数.

Ditto算法在原始的本地目标函数基础上增加了一个公平性正则项,从而优化每个客户端的更新方向. 另外,由于联邦学习中会存在恶意节点破坏全局模型的训练,为了保证算法的鲁棒性,Ditto通过超参数λ在个性化模型和全局模型之间进行权衡. 当全局模型的训练效果较好时,可以设置较大的λ值,个性化模型v_k将会更加接近全局模型w. 当全局模型被恶意节点攻击时,设置较小的λ值,能够使个性化模型v_k偏离受攻击的全局模型w. 通过动态调整λ的值,客户端能够在

全局模型和本地模型之间找到适合自己的个性化模型. 通过这种方法能够同时实现公平性和鲁棒性的保证.

在传统的联邦学习算法中为了加速模型的收敛速度和降低通信成本,通常会在客户端本地进行多个时间段的本地优化迭代,之后再上传服务器进行全局聚合. 然而,这种方法在数据异构性和设备异构性场景中会使本地模型更新方向之间的差距变大,从而阻碍模型的收敛,最终导致聚合后的模型会偏向本地损失最小的模型. 为了消除这类训练轮次造成的偏见,控制变量的方法经常被用于保证训练的公平性. 控制变量方法能够降低随机梯度的方差,减少模型之间的差异性,促进客户端性能之间的平均分配.

Karimireddy等人[52]提出了随机控制平均算法(Stochastic Controlled Averaging algorithm, SCAFFOLD),使用控制变量的方法修正本地更新中客户端之间的偏见,并验证了方法能够显著减少通信轮次且不会受到数据异构性的影响. 具体地,SCAFFOLD引入服务器和本地客户端的控制变量,用于估计服务器模型的更新方向和每个客户端的更新方向,然后利用这两个更新方向的差异来近似局部训练的偏移. 在每轮的客户端训练中计算客户端和服务端的差异,并增加了正则项 $c-c_i$ 优化每次本地更新的方向,从而减少客户端模型更新的偏见,使模型具有更快的收敛速度.

3.4 全局聚合更新

联邦学习的过程不仅包括了机器学习的算法,也包含了服务器与客户端之间的传输通信和本地训练完成后的全局聚合过程. 在联邦学习的全局聚合阶段,通过修改客户端上传下载的模型更新或全局聚合规则,使最终聚合得到的全局参数更新能够减少客户端之间的性能差异是另一种解决方案.

聚合权重是指每个客户端每轮上传的模型更新信息在全局聚合过程中所占的比重. 如果参与联邦学习的所有客户端数据独立同分布并且具有相同的计算设备时,服务器只需要以与全局目标相同的方式对上传的本地更新进行聚合. 然而,在实际场景中不同客户端之间通常存在着数据异构性和设备异构性,服务器端全局聚合规则的设计就变得非常重要. 根据客户端的本地更新情况,通过对聚合规则的修改,使聚合后的模型能够保证客户端之间性能的公平,从而提高联邦学习的公平性. 如果聚合规则设计得不合适,全局模型将会朝着某些本地客户端的更新方向偏离,并收敛到其本地目标的最优值点,造成联邦学习客户端之间的不公平.

Yurochkin等人[53]在概率联邦神经匹配(Probabilistic Federated Neural Matching, PFNM)中提出了模型神经网络参数的排列不变性问题,即对于不同客户端的本地更新,相同位置神经元的重要性可能并不相同,若只是简单进行聚合,会对模型的性能造成影响. PFNM算法是在客户端本地模型训练结束后,对模型的神经元进行重要性计算和重排,并将重要性相同的神经元进行匹配,然后再对本地模型的神经元进行聚合. PFNM相比FedAvg具有更好的性能和通信效率,但它只是在简单的体系结构上进行了测试,适用于简单的神经网络.

在之后的研究中,Wang等人[54]提出了联邦匹配平均(Federated Matched Averaging, FedMA)方法,将PFNM算法扩展到深度神经网络架构CNN和LSTM模型中. FedMA算法提出了按层构建全局共享模型的分层匹配方案. 具体地,服务器首先从客户端收集第一层的模型权重,并通过单层匹配的方法获取全局模型的第一层权重. 然后服务器会将收集到的权重广播给客户端,客户端继续对其数据集上的所有连续层进行训练,并保持联邦匹配层处于冻结状态. 随后不断重复此过程,直至最后一层. 服务器根据每个客户端数据集的类别比例对最后一层模型参数进行加权平均. 这种方法能够通过层级化的匹配机制实现客户端之间性能的优化. 然而,FedMA算法需要联邦学习的通信轮数与神经网络的层数相同,灵活性有待提高.

Zeng等人[55]提出了一种基于模型贡献的FedCav算法,将每轮客户端本地更新的贡献进行量化,并设计了一种新的全局损失函数用以迭代修正客户端模型训练的梯度下降过程. 通常情况下具有较高推理损失的本地数据可能有助于模型性能更好地优化. 另外,该方法能够根据客户端的本地历史训练统计数据来识别带有虚假损失的异常更新,从而保证了算法的公平性和鲁棒性.

Wang等人[56]针对不同用户数据集之间的分布差异以及网络状态不稳定带来的掉线问题,探索了联邦学习客户端之间性能差异产生的原因,并提出了联邦公平聚合(Federated Fair aVeraging, FedFV)方法,通过对联邦学习全局聚合规则的修改,减少客户端模型之间的差异. 具体地,首先使用顺差法来探测客户端之间的梯度冲突,然后通过修正梯度的方向和大小减少客户端本地更新方向之间的冲突. 这种方法能够提高模型的最终性能并实现公平性的保证.

全局聚合更新不需要重新设计客户端本地模型训练算法的目标函数,而是对全局聚合规则或客户端与服务器之间传输的模型更新信息进行修改. 然而,如何合理地修改不同客户端的模型更新和全局聚合规则,需要对每个客户端有很多的先验知识. 另外,模型更新

信息和全局聚合规则的修改对算法的精度和效率会有一定程度的影响. 因此,如何设计合理的全局聚合规则并在精度、效率和公平性之间做出很好的权衡是这类方法研究的核心.

本节梳理总结了基于平均分配理论联邦学习公平性的内涵、研究挑战和主要解决方案,其中代表性的方法总结如表2所示. 通过上面的分析可以看出,平均分配理论的研究目标是减小联邦学习客户端之间的收益差异,促进每个参与者最终获得的收益均衡,同时还需要在模型精度效率和公平性之间进行权衡. 具有代表性的方法是节点选择机制、本地模型优化和全局聚合更新. 然而,在实际情况中提高性能较差设备的重要程度来提升其模型性能,使所有设备具有相同性能的方法并不能解决联邦学习中所有的公平问题. 这类方法没有考虑到贡献度高的客户端,并会潜在地搁置高质量客户端的模型优化和性能,导致贡献高的客户端无法得到公平的对待,最终可能不愿进行数据共享和联邦学习. 因此,需要根据每个客户端在联邦学习中的实际贡献设计公平性机制. 我们在下节将详细介绍基于按劳分配理论联邦公平学习的相关研究.

表2　基于平均分配理论的联邦学习公平性常用方法总结

公平方法	解决方案	技术方法	工作特点
RBCS-F	节点公平选择	设计一个恒定的公平性参数作为长期公平性约束,以确保每个客户端的平均参与率不低于预期的保证率,以实现公平的FL客户端选择	设置准入阈值
Power-of-Choice	节点公平选择	设计了一个通信和计算高效的客户端选择框架,通过评估和选取偏向于具有更高本地损失的客户端,从而产生更快的误差收敛.	动态选择控制
AFL	本地模型优化	最大化联邦学习中最差客户端的性能,只要其他客户端的优化过程不增加性能最差客户端的损失,那么他们的模型性能就不会受到这种优化算法产生的负面影响	重新分配权重
q-FFL	本地模型优化	通过参数q进行权重再分配,使得具有较高损失的设备将会被分配更高的相对权重,从而能够获得更多的优化机会以此来减少模型之间的差异	重新分配权重
Ditto	本地模型优化	在原始的本地目标函数基础上增加了一个公平性正则项,从而优化每个客户端的更新方向	增加公平正则项
PFNM	全局聚合更新	客户端本地模型训练结束后,对模型的神经元进行重要性计算和重排,并将重要性相同的神经元进行匹配,然后再对本地模型的神经元进行聚合	修改聚合规则
FedFV	全局聚合更新	探测客户端之间的梯度冲突,然后通过修正梯度的方向和大小减少客户端本地更新方向之间的冲突,从而减少客户端模型之间的差异	修改模型更新

4　按劳分配理论

平均分配理论在联邦学习公平性研究中占了很大的比重,然而它并不总是合理的,不能解决联邦学习中所有公平性问题. 平均分配理论的联邦学习方法目标是每个客户端具有相同的收益而不管他们的实际贡献如何. 这类方法对在训练过程中贡献更多的客户端来说并不公平,这样会极大地损害具有高质量数据和计算能力的客户端参与联邦学习的过程. 另外,联邦学习中客户端之间是彼此陌生和不信任的,而全局模型是根据客户端自身上传的模型更新聚合形成的,这就不可避免地会存在自私或恶意破坏的客户端,他们可以通过上传虚假或恶意伪造的更新信息获取或破坏全局模型的更新过程. 例如,恶意客户端可以在训练过程中不提供任何数据,但是在每轮还是会获得全局更新的模型,这种现象被称为搭便车攻击[57]. 由于不管参与者如何进行操作都会接收到和其他参与者相同的模型,搭便车攻击在联邦学习的过程中会非常普遍. 恶意参与者的搭便车攻击最终会导致联邦学习面临着模型隐私泄露的问题,同时也会极大地损害高质量客户端参与联邦学习的积极性. 因此,需要根据每个客户端的实际贡献设计一种全新的联邦学习公平性机制为其分配相匹配的公平回报.

4.1　研究挑战

按劳分配理论的核心思想每个客户端模型的计算回报应该与他们在训练过程中的实际贡献相匹配. 从直观上看,拥有数量大和质量好数据集的客户端应该能够做出更高的贡献,贡献更多的客户端应该获得更好的回报,若两个客户端对联邦学习训练过程的贡献相同,则它们应该获得相同的回报. 为了达成这个目标,至少存在如下几个挑战需要得到有效解决.

4.1.1　如何评估客户端的实际贡献

在实际的联邦学习中,每个客户端可以实时生成和收集数据,这使客户端的本地数据集之间通常是非独立同分布,造成数据异构性的问题. 此外,客户端设备在计算和通信能力之间也存在差异,并且能够随时动态地参与联邦学习的训练过程,造成设备异构性的问题. 在联邦学习训练的过程,这种数据异构性和设备异构性导致每个客户端的实际贡献程度并不相同,有

必要对每个客户端的实际贡献进行评估. 另外,由于联邦学习隐私保护机制的设置,很多机器学习的贡献评估方法无法直接应用在联邦学习中. 因此,如何在联邦学习隐私保护的前提下,对所有参与客户端的贡献程度进行准确评估,是按劳分配理论研究的核心.

4.1.2 如何设计公平的贡献回报机制

在平均分配理论的研究中算法优化使所有参与者尽可能具有相同的回报,而不管他们在联邦学习中的实际贡献. 然而,平均分配理论使参与者即使上传较差的模型更新信息也会获得和其他客户端相同的分配结果. 如果每个人都得到平均的奖励,不管他们的贡献如何,具有高质量数据和计算能力的客户端可能没有动力参与数据共享和联邦学习的过程. 因此,需要设计公平的贡献回报机制使参与客户端收获的奖励与其实际贡献相匹配.

4.1.3 如何激励用户持续参与联邦学习

在用户参与联邦学习的过程中不可避免会存在设备计算和通信等资源的消耗,若用户在联邦学习中无法得到公平的对待和收益,那么他们参与联邦学习的积极性将会受到影响. 因此,不仅需要为每个参与的客户端设计与其实际贡献相匹配公平回报机制,同时也需要设计激励机制以鼓励更多的用户能够持续参与到联邦学习的过程中.

基于此,下文针对这些挑战分别介绍了目前按劳分配理论研究中的主要解决方案,包括贡献评估机制的设计、公平回报机制的设计以及激励机制的设计. 具体的技术分类如图5所示.

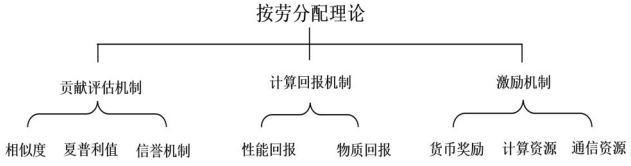


图5 按劳分配理论公平性的技术分类

4.2 贡献评估机制

贡献评估机制是对每个参与联邦学习的客户端在数据共享和模型训练过程中的实际贡献程度进行量化评估. 机器学习中所有的数据可以被集中访问和收集,并能够对数据进行集中质量评估. 而在联邦学习中由于数据隐私保护的设置,各方都难以访问其他客户端的原始数据,因此对他们的贡献程度进行评估更具挑战性. 目前对联邦学习客户端的贡献进行评估的机制尚未形成统一的评判标准,不同的研究者对评估机制有着不同的设计[58~61]. 我们将其归纳总结为数据贡献和模型贡献.

客户端本地的数据质量会对机器学习模型的性能造成很大的影响,因此不同客户端自身的数据质量能够反映他们在联邦学习中的数据贡献程度. 夏普利值(Shapley Value,SV)是一种计算每个参与者贡献的方法,起源于合作博弈论[62~64]. 夏普利值能够计算联盟中每个个体的平均边际贡献,保证利益分配中的公平,目前在机器学习[65~68]和联邦学习[69~72]中已被广泛应用. 夏普利值方法计算个体贡献程度的公式定义如下:

$$\varphi_k(S,v)=\sum_{s\in S/\{k\}}\omega(|s|)[v(s\cup k)-v(s)] \tag{15}$$

其中,$\varphi_k(v)$是第k个参与方在联盟S中的贡献度,$s\in S/\{k\}$是不包含第k个参与方后联盟S的所有非空子集,$\omega(|s|)$是加权的系数. 简单来说,夏普利值计算的是每个个体加入联盟后带来的边际利益.

夏普利值方法能够很好地评估个体的贡献,然而在联邦学习中应用时也存在着一些问题. 首先,夏普利值的计算过程需要计算个体对其他所有个体每个可能子集的平均贡献,在联邦学习的分布式场景中通信成本和计算成本会随着参与者数量或特征维度的增长而指数级增加. 因此,在联邦学习的数据评估中,研究者们提出相关方法来提高夏普利值的计算效率. Ghorbani等人[73]利用夏普利值来量化每个客户端对于联邦学习训练的贡献,提出了一种截断蒙特卡洛夏普利(Truncated Monte Carlo Shapley,TMC-Shapley)方法,通过使用随机抽样蒙特卡洛估计和截断每个排列抽样轮中不必要的效用评估来提高计算效率. Jia等人[74]提出使用组检验来加速夏普利值的计算过程. 这些方法侧重于减少每个个体在计算夏普利值过程中的排列抽样数. 然而,在联邦学习的过程中,这类方法在计算每一个客户端贡献的排列组合时,对应的联邦学习模型都需要从零开始训练. 这需要每个客户端产生大量的通信和计算开销,使这类方法不适合实际的联邦学习应用. 在此基础上,Liu等人[75]提出了导向截断梯度夏普利值(Guided Truncation Gradient Shapley,GTG-Shapley)方法,通过梯度聚合计算客户端的夏普利值,从而减少了计算量并且不用重新训练模型. 该方法还设计了一种截断的抽样方法,减少夏普利值计算中的排列组合数,进一步地提升了计算效率.

此外,夏普利值的计算过程没有考虑个体的计算顺序. 在联邦学习中,每个客户端的计算顺序可能会影响于它被用于训练的时间. 例如,为了确保收敛,模型更新会随着时间的推移而减少(例如,通过使用衰减的学习率). 因此,在学习过程结束时使用的数据源可能比以前使用的数据源更不实用. Wang等人[76]提出了一种适用于联邦学习的夏普利值计算方法. 具体地,夏普利值通过从每次训练迭代的局部模型更新中计算,这样不会产生额外的通信成本. 另外,该方法检查了学习

过程中每个客户端子集按照实际参与顺序所带来的性能提升,从而避免客户端的参与顺序对夏普利值的影响. Xu 等人[77]提出了一种余弦梯度夏普利值(Cosine Gradient Shapley Value,CGSV)方法用于计算客户端的贡献程度,使用客户端本地模型梯度之间的余弦相似度作为实际贡献,并通过理论证明得出余弦相似度能够有效而准确地近似夏普利值. 夏普利值在理想条件下能够有效地评估个体贡献程度,然而在联邦学习这种复杂场景中夏普利值的计算仍存在着问题亟须研究. Zhang 等人[78]通过实验分析发现即使同样的数据集和参与者,在不同的模型下不同参与者的夏普利值并不相同. 因此,夏普利值可能不完全适用于所有联邦学习场景中的客户端贡献评估.

基于数据贡献的评估研究都建立在客户端本地数据对训练贡献度相同的假设条件下,在实际联邦学习场景中这种假设很难得到保证. 例如,当训练数据存在异构性的情况下,每一个训练数据对模型的贡献程度都不相同. 另外,若在贡献评估的过程中不考虑客户端训练过程中的模型质量,那么客户端可以随意通过上传虚假或恶意伪造的参数更新获取全局模型的更新,从而进行搭便车攻击等恶意破坏行为,导致客户端之间的分配不公平. 因此,单纯以数据质量作为客户端贡献的评价标准,没有考虑到客户端在模型训练阶段的情况,不能完整地反映客户端的实际贡献. 为了更好地计算客户端的贡献度,客户端在训练过程中的模型贡献评估同样至关重要.

对模型贡献的评估常用方法是在每一轮中对比每个客户端的本地模型更新和全局模型更新的相似程度. 本地更新与全局更新越相似,表明该客户端的模型贡献程度更高. Chen 等人[79]将本地模型和全局模型的交叉熵损失函数(mutual cross-entropy loss)作为模型的贡献. Xu 等人[80]将本地模型和全局模型的余弦相似度作为模型的贡献. 另外,还有一种思路是客户端之间根据各自的模型进行互相评估. Lyu 等人[81]提出了一种客户端之间通过彼此生成的样本进行相互评价的方法. Pandey 等人[82]提出了一种参与客户端之间相互评估的方法. 具体地,客户端之间可以交换必要的信息,并使用本地数据集的相对准确率对其他参与者的贡献进行评估. 然而这些方法在一定程度上都与联邦学习隐私保护的本质相悖. Yan 等人[83]提出了一种基于注意力机制的联邦学习贡献评估(Federated Contribution Measurement,FedCM)方法,该方法每轮都会结合客户端本轮和上一轮的表现综合对其贡献程度进行评估更新,保证了贡献评估的实时性,该方法对数据数量和数据质量具有较高的敏感性. Nishio 等人[84]提出了一种基于逐步累计的贡献评估方法,该方法具有较高的计算和通信效率,适用于没有足够计算和通信预算的场景中. Zhao 等人[85]提出了一种基于强化学习的贡献评估方法. 具体地,客户端服务器计算其本地数据的参数梯度,并将梯度上传到中心服务器. 在聚合客户端梯度之前,中央服务器使用强化学习技术训练梯度的数据值估计器.

对于模型贡献评估来说,当前的方法通常假设客户端本地模型之间或本地模型和全局模型的相似度越高代表模型更有价值. 然而,在实际的联邦学习中这种假设可能并不总是成立的. 若参与的客户端寻求协同计算完成不同的任务,那么拥有互补知识数据的客户端可能比拥有相似数据的客户端更有价值.

4.3　公平回报机制

如何根据客户端的实际贡献为其分配相匹配的公平回报同样是一个重要的问题. 研究者们提出了不同的方法为客户端分配相应的回报. 根据回报形式的不同我们将其分为基于性能的回报和基于物质的回报.

基于性能的回报是根据客户端的贡献使他们获得相匹配的模型性能作为回报. 尽管考虑物质奖励作为回报在现实生活中是很常见的,但是由于联邦学习自身的机制设置,这类方法的实现可能并不容易. 因此,设计不需要物质奖励的回报机制同样重要,在联邦学习中通常将客户端最终模型的性能当作贡献的回报,贡献高的客户端将会获得到更高的模型性能.

Lyu 等人[81]提出了一个去中心化的公平和隐私保护深度学习(Fair and Privacy-Preserving Deep Learning,FPPDL)框架,每个参与者根据其在联邦学习中的贡献获得最终的模型性能. FPPDL 的核心思想是客户端可以将信息提供给其他参与者来赚取积分,然后他们可以使用赚取的积分从其他参与者获取信息. 具体地,在 FPPDL 中所有客户端之间通过参与者之间的相互评价来维护一个共识的信誉体系,每个参与者根据自身的信誉值能够获得相应数量的积分,然后参与者可以使用他们的积分与其他参与者进行交易. 与传统的联邦学习框架相比,贡献更多的客户端会有更高的信誉值和积分,并且每个参与者获得的回报与其相应的贡献成正比.

Xu 等人[77]将客户端每轮从服务器下载的梯度作为贡献回报,通过稀疏梯度的方法,根据每个客户端上传的模型梯度信息在整体中所占的比例,决定其从服务器端下载的梯度质量. 不同的梯度质量最终会影响每个客户端的性能表现,使模型贡献更多的客户端会获得性能更好的模型.

Zhang 等人[78]提出了层次公平联邦学习(Hierarchically Fair Federated Learning,HFFL)框架,将客户端按

照其自身的贡献划分级别，每个级别训练符合自身等级的联邦学习模型，从而确保客户端之间的公平．具体地，HFFL首先根据客户端的贡献程度（如数据质量、数据数量等）将客户端划分为不同的级别，每个级别都会训练一个联邦学习模型．当训练较低级别的联邦学习模型时，高级别的客户端只需提供与低级别的客户端数量相同的数据．当在训练较高级别的联邦学习模型时，低级别的客户端需要提供所有的本地数据．通过这种方式，高级别训练得到的模型性能会高于低级别的性能，从而保证了按劳分配的公平．然而，HFFL中仍然存在着一些问题．HFFL对等级的划分依靠客户端数据的数量，并没有综合考虑客户端的全部贡献，在同一级别中客户端的质量可能也并不相同．

基于物质的回报主要是为客户端分配相应的货币奖励，通常是与激励机制相结合使用．我们将在下文中进行详细探讨．

4.4 激励机制

激励机制是在此基础上通过设计方案鼓励更多的客户端积极参与到联邦学习的过程中，从而实现联邦学习的高效协同计算和可持续发展[86~89]．在实际场景中当客户端参与联邦学习进行协同计算时，不可避免地会出现通信和计算开销等设备资源的消耗．此外，联邦学习的框架仍面临着各种安全风险，许多数据拥有者并不会积极主动地参与进来，特别是当他们拥有的是敏感和隐私数据时．因此，在保证用户隐私的前提下，需要通过设计激励机制鼓励数据拥有者参与到联邦学习的过程中．为了保证联邦学习的公平性，激励机制的设计通常会为每个参与的客户端分配与其实际贡献相匹配的回报．由于联邦学习的隐私保护机制，客户端之间不会知道彼此的真实数据，量化每个客户端的数据价值和模型价值是很困难的，对每个客户端的贡献进行建模也很困难，这使现有的关于激励机制设计的方法不能直接应用在联邦学习上．关于贡献评估和公平回报的方法我们在前面已经进行了总结，在这一节中我们重点从公平性的角度探讨激励机制中的方法．

在激励机制的设计中，客户端在使用本地数据集参与联邦学习时，通常会获得货币奖励或资源奖励作为贡献回报．该机制通过这种方法激励更多的用户参与联邦学习．典型的激励机制框架如图6所示．目前联邦学习公平性的激励机制奖励形式主要包括货币奖励[90]、计算资源[91]和通信资源[92]的奖励等．激励机制中的奖励设置可以分为两种方式：第一种是在联邦学习训练之前对客户端进行奖励，从而激励高质量的用户参与到联邦学习的过程中；第二种是在联邦学习训练完成后为其分配相应的奖励，保证客户端之间的分配公平，促进用户持续地参与联邦学习．

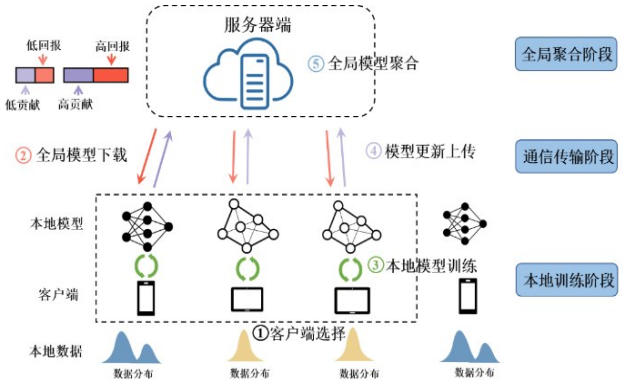


图6 激励机制框架

联邦学习训练前奖励的目标是激励更多的高质量用户进行数据共享并参与到联邦学习过程中，从而协同计算获得更好的全局模型．奖励可以根据客户端能够贡献资源的质量和数量决定，也可以通过建立信誉评价体系由客户端之间的投票决定[93]．训练后奖励考虑了客户端在联邦学习过程中的实际贡献，并为其分配与贡献相匹配的奖励[94]．它的目标是使联邦学习的分配更加公平，从而促进客户端能够持续地参与联邦学习．

客户端本地的数据质量会对机器学习模型的性能造成很大的影响．因此不同客户端的数据质量能够反映他们在联邦学习中的数据贡献程度，从而根据贡献程度设计相应的激励机制．Song等人[95]提出了一种基于夏普利值的评价指标用以评估不同客户在联邦学习中的贡献．Sim等人[96]设计了一种基于客户端训练数据贡献的激励机制，使用夏普利值计算客户端的贡献，并通过为每个客户端分配定制模型作为奖励．为了吸引拥有高质量数据的客户端参与联邦学习，Yu等人[97]提出了FLI框架，在该框架中持续贡献高质量数据并且长期未得到合理奖励的客户端将在之后获得更高份额的收入，从而激励数据拥有者能够在联邦学习中持续贡献高质量的数据．在设计奖励的同时也需要考虑到奖励的实时性．Yu等人[98]在提出了一种能够动态分配给定预算的收益分成方案，解决了支付延迟造成参与者的积极性降低的问题．

拍卖机制和契约理论也常被应用于联邦学习激励机制的设计中[99]．针对联邦学习中数据异构性和设备异构性的问题，Zeng等人[100]提出了一种多维联邦学习激励机制，通过博弈论推导出每个客户端的最优分配策略．在模型训练和聚合过程中，服务器端会根据这些策略选择最优的客户端参与．Kang等人[101]设计了一种基于契约理论的激励机制，以鼓励拥有高质量数据的客户端加入联邦学习过程，并平衡任务发布者和数据拥有者之间的效用．这种机制能够确保联邦学习的公平性和有效性．Lim等人[102]提出了一种分层激励机制

框架,利用契约理论建立客户端之间的激励机制. 该机制能够在信息不对称的情况下,利用博弈论的方法根据客户端的边际贡献为其分配奖励,激励不同的客户端持续提供高质量的数据.

联邦学习中设备的异构性使客户端之间的训练受到不同的资源约束,包括带宽、存储和能量,如何设计激励机制为不同客户之间分配公平的资源变得至关重要. 基于资源分配的激励也是激励机制设计的重要方向,包括计算资源和通信资源的奖励.

在计算资源奖励方面,当计算资源有限时,为了使客户端积极参与协同计算并贡献计算资源,可以通过将预算分配给客户端的方法激励更多参与者. Sarikaya等人[103]首先分析了设备异构性对联邦学习的影响,并在此基础上提出了一个斯塔克尔伯格(Stackelberg)博弈模型,通过优化客户端之间的计算资源分配策略以及参数服务器的预算分配来提高联邦学习的性能,通过激励机来平衡客户端在每次迭代过程中的时间延迟. Khan等人[104]也采用了Stackelberg博弈的思想来设计激励机制,以激励用户积极地参与到联邦学习中. 另外,强化学习也被用在激励机制的设计中. Zhan等人[105]提出了一种基于深度强化学习(Deep Reinforcement Learning,DRL)的方法,能够在动态网络环境下找到模型训练时间和参数服务器预算分配之间的最佳权衡. 由于联邦学习算法的复杂性和动态学习环境的限制,传统的理论分析方法可能难以寻求最优解. 相比之下,DRL方法可以利用以前训练过程中积累的经验来改进当前策略,从而使当前策略接近最优解,为联邦学习公平性的研究提供了一种新的思路[106, 107].

在通信资源方面,Pandey等人[82]从模型参数交换期间的通信效率成本考虑,提出了一个新的众包框架,采用两阶段Stackelberg博弈方法求解双方利益最大化的优化问题. 此外,他们还为参与联邦学习的客户端提供准入控制方案,以确保联邦学习的性能和公平性. Le等人[108]设计了一种激励机制,将服务器端和客户端之间的关系定义为一种拍卖机制,其中服务器端作为拍卖商,客户端则是卖家. 客户端使用最优的通信资源分配策略,在满足联邦学习延迟约束的同时,使计算消耗最小化. 在服务器端拍卖机制博弈中,客户端选择问题被描述为效用最大化问题,并提出了一种原始-对偶贪婪算法来求解这个NP难问题.

区块链因为其去中心化、公开透明和防篡改的特点也经常用在激励机制的设计中[109~113]. Kim等人[114]提出联邦学习和区块链结合的BlockFL框架,采用区块链网络共享客户端的模型参数,并且为每个客户端分配的奖励与其参与联邦学习的训练数据量相匹配. Weng等人[115]同样使用区块链网络来共享模型参数,并设计基于训练数据量和用户信誉值的奖励分配方案,保证了激励的相容性和联邦学习的公平性. Zhang等人[116]使用客户端的数据量和数据类别之间的质心距离来衡量用户的贡献,并基于此为其分发区块链的代币作为奖励,激励客户端参与联邦学习.

本节梳理总结了基于按劳分配理论联邦学习公平性的内涵、研究挑战和主要解决方案,其中代表性的方法见表3. 按劳分配理论的研究目标是参与客户端在联邦学习中获得的回报要与他的实际贡献相匹配. 具有代表性的方法包括贡献评估机制、公平回报机制和激励机制的设计. 然而,在实际情况中按劳分配理论的方法也并不能解决联邦学习中所有的公平问题,因为这类方法会使性能好的客户端越来越好,导致客户端之间的差距越来越大. 另外,在当前的贡献评估和回报机

表3　基于按劳分配理论的联邦学习公平性常用方法总结

公平方法	解决方案	技术方法	工作特点
CSGV	贡献评估机制	提出了一种余弦梯度夏普利值(CGSV)方法计算客户端的贡献程度,使用客户端本地模型梯度之间的余弦相似度作为实际贡献,并通过理论证明得出余弦相似度能够有效而准确地近似夏普利值	模型梯度相似度
GTG-Shapley	贡献评估机制	提出了GTG-Shapley方法通过梯度聚合计算客户端的夏普利值,从而减少了计算量并且不用重新训练模型	梯度聚合计算夏普利值
FLI	公平回报机制	提出了FLI框架,在该框架中持续贡献高质量数据并且使长期未得到合理奖励的客户端在之后获得更高份额的收入,从而激励数据拥有者能够在联邦学习中持续贡献高质量的数据	修改分配规则
HFFL	公平回报机制	将客户端按照其自身贡献划分级别,每个级别训练符合自身等级的联邦学习模型,从而确保客户端之间的公平	按等级划分参与方
Fmore	激励机制	提出了一种多维的联邦学习激励机制,利用博弈论推导出每个客户端的最优分配策略,服务器端会根据推导出的策略选择最优的客户端参与模型训练和聚合的过程	利用博弈论推导策略最优分配
Deepchain	激励机制	使用区块链网络来共享模型参数,并设计基于训练数据量和用户信誉值的奖励分配方案,保证了激励的相容性和联邦学习的公平性	利用区块链分发代币作为奖励

制的研究中,初期的评估效果对于整体效果的影响过大.若上一轮在评估中没有获得很好的评价,那么在下一轮训练阶段将会处于劣势,这种劣势将持续下去并逐轮累计.因此,在未来的研究中需要结合实际应用场景在平均分配理论和按劳分配理论中做出平衡,提出更加符合联邦学习实际应用公平性定义.

5 联邦学习公平性学术资源总结

联邦学习公平性的研究目前仍处于起步阶段,前文已经对联邦学习公平性的内涵、研究挑战和解决方案进行了详细探讨.本节将重点对研究涉及的常用数据集、实验场景设置和公平评价指标进行全面归纳总结.

5.1 常用数据集

联邦学习公平性在不同的应用场景中有着不同的定义和内涵.当前联邦公平学习研究中的常用数据集如表4所示.这些数据集涵盖了不同的领域,并列出了使用这些数据集的代表性文献,以探究联邦学习公平性在不同数据集中的应用情况.

表4 联邦学习公平性常用数据集总结

数据集名称	描述信息	训练样本	测试样本	特征数量	类别数量	代表文献
MNIST	手写字符识别数据集	60 000	10 000	784	10	文献[25,33,58,75] 文献[77,78,90,95]
FEMINIST	手写字符识别数据集	805 263		784	62	文献[18,19,25,27,51]
EMNIST	手写字符识别数据集	814 255		784	62	文献[19,25,52]
Fashion-MINIST	时尚物品图像识别	60 000	10 000	784	10	文献[19,35,48] 文献[59,78,90]
CIFAR-10	图像分类	50 000	10 000	1 024	10	文献[33,51,56,77]
CIFAR-100	图像分类	50 000	10 000	1 024	100	文献[19,25]
Sent140	文本情感分析	1 600 498		—	—	文献[18,48,50]
Shakespeare	采集自莎士比亚作品全集预测下一个字符的语言模型数据集	4 226 158		—	—	文献[25,27,48,50,51]
Adult	人口普查数据库中提取的数据	32 561	16 281	123	2	文献[25,35,49,78]
Camelyon17	医学病理图像数据集	—	—	—	—	文献[117,118]

MNIST:MNIST数据集(Mixed National Institute of Standards and Technology database)是美国国家标准与技术研究院收集整理的大型手写数字数据库,其中包含60 000个示例的训练集以及10 000个示例的测试集.每张图片有28×28=784个像素,表达了0~9这10个数字中的一个.

FEMINIST[119]:手写字符识别数据集,其中包括62种不同的字符类别(10种数字,26种小写,26种大写)的像素图片,每张图片有28×28=784个像素,样本数为805,263,可使用官方给出的代码按3 500个客户进行非独立同分布划分.

EMNIST[120]:手写字符识别数据集,是MNIST数据集的扩展版.其中包括62种不同的字符类别(10种数字,26种小写,26种大写)的像素图片,每张图片有28×28=784个像素,样本数为814 255.

Fashion-MINIST[121]:时尚物品图像识别数据集,包含10种类别不同的时尚物品,其中包含60 000个示例的训练集和10 000个示例的测试集,每张图片具有28×28=784个像素.

CIFAR-10[122]:图像分类的数据集,包括10种类别的彩色图片(包括人、动物、花、昆虫等),每种类别都有6 000张图片.其中包含50 000个示例的训练集和10 000个示例的测试集,每张图片具有32×32=1024个像素,在联邦学习中可以根据自行设计的场景进行划分.

CIFAR-100:图像分类的数据集,包括100种类别的彩色图片(包括人、动物、花、昆虫等),每种类别都有600张图片.其中包含50 000个示例的训练集和10 000个示例的测试集,每张图片具有32×32=1024个像素,在联邦学习中可以根据自行设计的场景进行划分.

Sent140:一个由许多推文组成的文本情感分析数据集,其中每条推文都可以提取为积极或消极的情感.每个设备都是不同的twitter用户,共1 600 498个样本.

Shakespeare:从莎士比亚的作品全集中采集而来,用于预测下一个字符语言模型的数据集.其中每个用户是作品中的一个角色,共有4 226 158个样本.

Adult:Adult数据集是从1994年美国的人口普查数据库中提取的数据,其中具有48 842条记录,每条记录包含年龄、职业、受教育程度、种族、性别、婚姻状况、出生地、每周工作时长等属性,这些属性有连续的和离散的.该数据集可以研究性别或种族对收入的影响,在公平性的研究中扮演重要角色.

Camelyon17：一个具有真实多中心数据分布的医学图像数据集. 该数据集的主要研究对象是乳腺癌,旨在促进计算机视觉和机器学习在医学影像领域的应用,以提高乳腺癌的早期诊断和治疗效果. 由于其真实多中心数据分布的特点,该数据集在联邦学习研究中也被广泛采用.

通过对联邦学习公平性中常用的数据集进行总结分析发现,大部分的文献选择使用图像数据集 MNIST, CIFAR-10, Fashion-MNIST, FEMNIST 和文字识别数据集 Shakespeare 等作为其评估的主要数据来源. 另外,通过对上述文献所应用的数据集进行规律总结,可以看出联邦学习公平性相比其他研究方向对数据集的使用更常见,其主要原因是联邦学习公平性研究中实验场景的设置通常是将真实数据集划分为多个较小的子集作为相同数量客户端的本地数据集,并通过对真实数据集不同的划分策略模拟非独立同分布的不同情况. 与直接使用真实的联邦数据集相比,数据集划分策略有以下优点:

(1)真实联邦数据需要收集所有参与方的本地数据,由于联邦学习的隐私保护特性,适用于训练的真实联邦数据集通常难以获取. 通过对现有广泛使用的公共数据集进行划分更具有灵活性,这些数据集已经有大量集中训练的知识作为参考.

(2)真实联邦数据集的数据资源通常是固定的,但数据集划分策略可以通过划分不同数据量的子集模拟不同数量的客户端以及不同的应用场景.

(3)真实联邦数据集中数据之间的不平衡程度通常很难评估,但通过数据集划分的策略可以很容易地量化和控制数据之间的不平衡程度.

同时,我们还能发现当前很多联邦学习公平性的方法在图像和文本数据上都进行了实验具有一定的通用性. 所以后续的研究可以多结合实际的场景和数据集进行联邦学习公平性的研究,推进联邦学习真正地在现实中落地应用.

5.2　实验场景设置

机器学习实验场景中由于数据可以被收集并存储在中心服务器,模型训练的过程可以获取所有数据的整体信息. 但是在联邦学习中,数据通过不同客户端收集并仅存储在本地,不同设备存在着通信能力和计算能力等差异,导致客户端之间存在着数据异构性和模型异构性. 另外,由于联邦学习的隐私保护设计,客户端之间不能访问彼此的私有数据,导致很多机器学习的算法无法直接应用在联邦学习中. 本节将对当前联邦学习公平性研究中常见的实验场景设置进行归纳梳理.

用户参与联邦学习的初衷主要是自身没有足够的数据去训练高质量的本地模型,希望通过与其他用户的协作计算获得性能更佳的模型. 对于那些自身拥有高质量数据的用户来说,在异构场景中参与联邦学习得到全局共享模型的性能可能不如他们自己训练的局部模型的性能,传统的联邦学习方法无法保证客户端之间的公平分配,最终用户可能不愿参与数据共享和联邦学习. 然而,联邦学习需要拥有高质量数据的用户持续参与联邦学习训练. 因此,如何应对联邦学习中存在的异构性问题是联邦学习公平性研究中亟须解决的问题.

为了研究联邦学习中的异构性,有必要先对异构场景进行划分,联邦学习中的异构性主要包括:

(1)设备异构性. 联邦学习中的客户端设备在存储、计算和通信能力方面都存在着差异,这会导致每个客户端本地模型训练时间的不同. 在每轮的训练过程中需要等待最慢的客户端训练结束才能进行上传聚合,这会严重影响联邦学习的效率,对训练快的客户端也是不公平. 若规定时刻上传,服务器无论是直接舍弃超时客户端的本地更新,还是将客户端未完成的本地更新直接上传,都会对最终全局聚合模型的更新造成很大影响. 设备异构性是影响联邦学习公平性的重要因素.

(2)数据异构性. 联邦学习中每个客户端的本地数据是通过自身收集得到,导致参与训练客户端之间的数据虽然独立分布但不服从同一采样方法(Non-Idependently and Identically Distributed, Non-IID,简称非独立同分布)[123~127]. 在联邦学习的过程中,每个客户端都是通过优化自身本地数据样本的期望损失训练模型. 在数据非独立分布的条件下,不同客户端之间的数据差异较大,本地模型之间的训练方向同样会相差很大,偏离全局模型的优化方向,最终导致联邦学习模型精度的严重下降以及客户端之间的公平性无法保证.

(3)模型异构性. 由于各个客户端设备异构性和数据异构性,本地训练得到模型同样存在着异构性. 联邦学习最初的目标是所有客户端通过聚合更新得到一个全局模型. 然而由于现实中客户端设备的异构性以及数据在客户端之间的非独立同分布,联邦学习很难聚合得到一个适合所有客户端的全局模型,客户端本地模型之间也会存在着很大的差异.

目前联邦学习公平性的研究主要针对的是数据异构性的问题,即如何在非独立同分布的情况下保证客户端之间的公平. 联邦学习公平性研究常见的数据划分策略有均衡分布、样本数量不平衡、基于数量的标签分布不平衡、基于分布的标签分布不平衡、基于噪声的特征分布不平衡和数据内在的特征不平衡[123, 124]. 表5对其进行了详细的总结. 通过不同的数据划分策略模

拟可能存在的非独立同分布场景,不管是哪种情况的非独立同分布,每个设备中的数据类别都不完备,不能代表全局的数据分布.下面将对不同的划分策略进行详细的描述.

均衡分布中每个客户端都有相同数量的样本,所有类样本的分布也相同.

样本数量不平衡场景中,本地数据集D_k的大小因客户端而异.可以使用狄利克雷分布或对数正态分布$Log-N\left(0,\omega^2\right)$为每个客户端分配不同的样本数,而对不同的类样本保持相同的分布.

基于数量的标签不平衡中每个客户端只有特定数量的样本类,标签的种类是固定的.具体的,假设每个客户端的数据样本只有C个不同的标签.首先随机分配C个不同的标签id给每个客户端.然后将每个标签的样本随机平均地分配给拥有该标签的客户端.通过这种方法每个用户的标签数量是固定的,不同用户的样本之间没有重叠.

基于分布的标签不平衡中每个标签都根据狄利克雷分布给每个用户分配一定比例的该标签的样本.具体的有,平衡的狄利克雷分类(balanced Dirichlet partition)每个客户端具有相同数量的样本,而每个客户端的类分布遵循狄利克雷分布.不平衡的狄利克雷分类(unbalanced Dirichlet partition)客户端的样本数量来自对数正态分布,而每个客户端的类分布服从狄利克雷分布.异质的狄利克雷分类(hetero Dirichlet partition)数据点的数量和类的比例不平衡.首先将样本划分为j个客户端,然后通过采样$p_i\sim \mathrm{Dir}_j(\alpha)$将第$i$类的样本以$p_{\{i,j\}}$的比例分配给相应的本地客户端.

基于噪声的特征不平衡中不同客户端的样本特征具有不同的高斯噪声水平.首先将整个数据集随机平均分成多个部分作为客户端的本地数据集,然后对每个客户端的本地数据集添加不同程度的高斯噪声,实现不同的特征分布.

数据内在的特征不平衡是由于数据集可能自身就存在着特征分布不平衡,例如在手写字符识别中,不同作者的字迹并不相同,因此根据不同作者将数据集划分成不同的子集作为客户端的本地数据集,这样不同客户端之间便有了天然的特征分布不平衡.

表5 联邦学习公平性常用实验场景设置总结

实验场景设置	公平性标准	主要描述信息	代表文献
均衡分布	平均、按劳	每个客户机被随机分配给所有类的统一分布,每个类的标量平衡	文献[19,27,55,75,77]
样本数量不平衡	平均、按劳	将数据集随机分成不同比例的数据大小.每个客户端的分布相同,但是样本数量明显不同	文献[75,77,78,95,123]
基于数量的标签不平衡	平均、按劳	每个客户端只有特定数量的样本类,标签的种类是固定的,不同用户的样本之间没有重叠	文献[28,51,55,56,75]
基于分布的标签不平衡	平均、按劳	每个标签都根据狄利克雷分布给每个用户分配一定比例的该标签的样本	文献[55,75,77,95,123]
基于噪声的特征不平衡	平均、按劳	通过对每个客户端的本地数据集添加不同程度的噪声,实现不同的特征分布	文献[28,75,95,123]
数据内在的特征不平衡	平均、按劳	数据集可能自身就存在着特征分布不平衡,造成不同客户端之间存在数据内在特征的不平衡	文献[18,19,25,27,51]

通过对联邦学习公平性中实验场景设置的总结分析发现,目前常用的实验设置包括均衡分布、标签分布不平衡、特征分布不平衡和样本数量不平衡.这些设置通过不同的数据划分策略,尽可能全面地模拟实际场景.不同的Non-IID分布对联邦学习算法的性能和公平性有很大的影响.另外,我们还可以通过在实验场景中使用不同的划分方法来探索更多的数据分布对联邦学习公平性的影响,并针对不同的场景探索不同的联邦学习方法来提高公平性和准确性.

5.3 公平评价指标

联邦学习的公平性在不同的应用场景中有着不同需求和定义,对于公平性的评价标准和指标也存在着不同.在本节中我们将对常用的联邦学习公平性评价指标进行归纳总结,具体如表6所示.

5.3.1 平均分配常用指标

平均理论的主要思想是减少客户端之间的差异,使用户具有相同的收益.在联邦学习中客户端的收益可以通过精度和效率体现.因此,主要的公平评价指标有平均准确率、标准差和风险差.

平均准确率(Average Accuracy):每个类别准确度的平均值,反映联邦学习中客户端模型的平均性能.

$$\mu=\frac{1}{N}\sum_{k=1}^{N}X_k \tag{16}$$

标准差(Standard Deviation):客户端之间的标准差通常用于衡量算法在不同设备上的性能变化,反映客户端之间模型性能的差异程度,在平均理论中能够衡量算法的公平程度.

$$\sigma=\sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(X_k-\mu\right)^2} \tag{17}$$

表6　联邦学习公平性常用实验场景设置总结

公平指标	公平标准	计算公式	描述信息	代表文献
平均准确率	平均、按劳	$\mu=\frac{1}{N}\sum_{k=1}^{N}X_k$	每个类别准确度的平均值,反映联邦学习客户端模型的平均性能	文献[18,19,35,48,49]
标准差	平均	$\sigma=\sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(X_k-\mu\right)^2}$	衡量算法在不同设备上的性能变化,反映组内个体间的离散程度	文献[19,25,48,51,56]
风险差	平均	$\left\lvert P\left(\hat{Y}=1\lvert S=1\right)-P\left(\hat{Y}=1\lvert S=0\right)\right\rvert$	衡量对敏感群体和非敏感群体的阳性预测值之间的差异	文献[49]
训练时间	平均、按劳	$T=T_e-T_s$	联邦学习过程花费的总时间,反映不同联邦学习方法的计算效率	文献[33,75,95]
余弦相似度	按劳	$\mathrm{CS}=\frac{\sum_{k=1}^{N}\left(x_k\times y_k\right)}{\sqrt{\sum_{k=1}^{N}x_k^2}\times\sqrt{\sum_{k=1}^{N}y_k^2}}$	通过向量空间中两个向量夹角的余弦值衡量两个变量之间的差异	文献[75,77,95]
欧氏距离	按劳	$\mathrm{ED}=\sqrt{\sum_{k=1}^{N}\lvert x_k-y_k\rvert^2}$	通过计算向量之间的绝对距离衡量两个变量之间的差异	文献[75,95]
最大差异度量	按劳	$\mathrm{MD}=\max\lvert x_k-y_k\rvert$	测量客户端在参与和不参与联邦学习之间的性能差异的最大百分比	文献[75,95]
皮尔森相关系数	按劳	$\mathrm{PCC}=\frac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$	衡量客户端实际贡献和贡献回报之间的相关程度	文献[35,77]
Jain公平指数	按劳	$\mathrm{JFI}=\frac{\left(\sum_{k=1}^{N}\frac{x_k}{z_k}\right)^2}{N\times\sum_{k=1}^{N}\left(\frac{x_k}{z_k}\right)^2}$	衡量用户是否获得了系统资源的公平共享	文献[33]

风险差(Risk Difference):衡量敏感群体和非敏感群体阳性预测值之间的差异,用于量化群体公平性.

$$RD(f)=\left\lvert P\left(\hat{Y}=1\lvert S=1\right)-P\left(\hat{Y}=1\lvert S=0\right)\right\rvert \tag{18}$$

其中,$\hat{Y}$是f的预测值,S表示是否是敏感属性. 公平性方法要求模型的预测结果独立于敏感属性S. 因此,风险差值越低表明该方法更加公平.

5.3.2　按劳分配常用指标

按劳分配理论的主要思想每个客户端模型的收益应该与他们在训练过程中的实际贡献相匹配. 因此,主要的公平评价指标有联邦学习的时间、欧氏距离、余弦相似度、最大差异度量、皮尔森相关系数和Jain公平指数.

联邦学习的时间(Time,T):联邦学习过程花费的总时间,反映联邦学习方法的效率.

$$T=T_e-T_s \tag{19}$$

其中,T_s和T_e分别是联邦学习训练开始和结束的时刻.

欧氏距离(Euclidean Distance,ED):欧氏距离用于计算多维空间中不同向量之间的绝对距离.

$$\mathrm{ED}=\sqrt{\sum_{k=1}^{N}\lvert x_k-y_k\rvert^2} \tag{20}$$

在联邦学习公平性研究中,欧氏距离常被用以计算客户端之间的相似度或客户端与服务器端之间的相似度. 欧氏距离的计算存在着量纲问题,在使用时需要保证各维度指标在相同的刻度级别.

余弦相似度(Cosine Similarity,CS):通过向量空间中两个向量夹角的余弦值衡量两个变量之间的差异.

$$\mathrm{CS}=\frac{\sum_{k=1}^{N}\left(x_k\times y_k\right)}{\sqrt{\sum_{k=1}^{N}x_k^2}\times\sqrt{\sum_{k=1}^{N}y_k^2}} \tag{21}$$

余弦相似度同样常被用以计算客户端之间的相似度或客户端与服务器端之间的相似度. 欧式距离体现数值上的差异,而余弦距离体现方向上的差异.

最大差异度量(Maximum Difference,MD):表示任意坐标维度上两个向量的最大差值. 与欧氏距离和余弦距离的通用性不同,最大差异度量通常用于特定的场景中.

$$\mathrm{MD}=\max\lvert x_k-y_k\rvert \tag{22}$$

其中,x_k和y_k可以分别表示客户端k参与和不参与联邦学习所获得模型的性能,从而计算模型性能之间差异的最大百分比.

皮尔森相关系数(Pearson Correlation Coefficient,PCC):皮尔森相关系数用于计算两个变量之间的相关程度.

$$\mathrm{PCC}=\frac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}} \tag{23}$$

在联邦学习公平性研究中,X和Y通常分别表示用户的实际贡献和贡献回报,从而反映按劳分配方法的公平程度. 皮尔森相关系数取值范围为[-1,1],接近0

的变量被称为无相关性,接近1或者-1被称为具有强相关性.

Jain公平指数(Jain's Fairness Index,JFI):Jain公平指数最初用于网络工程中的公平性度量,以确定用户或应用程序是否获得了系统资源的公平共享;后来研究者将其应用在联邦学习公平性研究中用于度量按劳分配方法的公平程度.

$$\mathrm{JFI}=\frac{\left(\sum_{k=1}^{N}\frac{x_k}{z_k}\right)^2}{N\times\sum_{k=1}^{N}\left(\frac{x_k}{z_k}\right)^2} \tag{23}$$

其中,x_k/z_k是一个关于客户端k的规范化值,它可以表示与联邦学习方案如何处理k有关的任何特征.例如,x_k和z_k可以分别表示用户的实际贡献和贡献回报,从而反映它们之间的相关程度.较高的JFI值意味着按劳分配的方法更加公平.

通过对联邦学习公平性中常用的公平评价指标总结发现,联邦学习模型的精度和效率(即平均准确率和训练时间)是两种理论都会关注的指标.除此之外,平均分配理论和按劳分配理论所使用的评价标准和公平指标并不相同.在平均分配理论的研究中常用的公平评价指标是标准差和风险差,其主要原因是平均分配理论需要衡量客户端模型之间的收益,并通过算法的设计减少收益的差异.在按劳分配理论的研究中,需要对客户端的贡献进行评估,区别客户端之间的贡献差异,并根据贡献程度的不同分配相应的公平回报.因此,客户端本地模型之间或客户端本地模型与全局模型的距离度量(如余弦相似度、欧氏距离和最大差异度量等)常被用作贡献评估的衡量标准.在按劳分配理论还需要为每个客户端分配与其贡献相匹配的奖励,因此,联邦学习中客户端获得奖励和实际贡献之间的相关系数(如皮尔森相关系数和Jain公平指数等)常用于计算奖励和贡献的相关程度,从而反映按劳分配方法的公平程度.

6 未来研究展望

联邦学习的公平性直接影响着用户参与数据共享和联邦学习的积极性和持续性.虽然联邦学习公平性的研究逐渐受到了关注,但总体而言,相关研究尚处于起步阶段,仍存在着如下亟须解决的挑战和研究难点.

6.1 场景适配的公平性度量

通过对平均理论和按劳分配理论的总结分析,我们发现它们仍存在着问题,不能解决联邦学习公平性的所有问题.在实际情况中,每个客户端参与联邦学习的动机可能并不相同,导致联邦学习不公平的来源种类多且复杂.参与联邦学习的客户端之间通常存在着设备异构性和数据异构性,不同的异构场景中不公平的影响程度也不相同.另外,公平性是相对的,针对不同的应用场景公平性有着不同的定义,例如平均分配理论和按劳分配理论.在不同的公平定义中,对公平性的评判标准是不同的,使用不同的公平评价指标去评估联邦学习学习公平性方法的性能往往会得到不同甚至是相反的评判结果.现阶段联邦学习公平性的定义和研究还无法包含所有的公平性场景,不同理论还未形成统一完善的公平性评价标准.

因此,如何针对性地选择适合具体任务场景的公平性度量是联邦学习公平性研究的前提.在未来的研究中,首先需要明确公平性所适配的场景,进一步地探索联邦学习中的公平性内涵,明确公平性度量在联邦场景中的泛化性,结合实际应用场景在平均分配理论和按劳分配理论中做出平衡,提出场景适配的公平性度量方法.

6.2 联邦学习的动态公平性

当前联邦学习公平性的研究场景通常是假设所有客户端同时同步参与联邦学习的过程,并由此对客户端之间的公平性进行优化,没有根据实际过程而动态变化[128].造成不同理论的公平性研究都还存在着许多问题,平均分配中的常见方法是通过提高收益较差设备的重要程度,这类方法没有考虑到贡献度高的客户端,并会潜在地搁置高质量客户端的模型优化和收益,导致贡献高的客户端无法得到公平的对待,最终可能不愿进行数据共享和联邦学习.按劳分配中初期的评估效果对于整体效果的影响过大.若上一轮在评估中没有获得很好的评价,那么在下一轮训练阶段将会处于劣势,这种劣势将持续下去并逐轮累计,导致客户端之间的收益差距越来越大.

实际中更常见的场景是参与方加入联邦学习有先后次序之分.未来需要研究这些动态环境下的联邦学习公平性方案,比如联邦已有部分固定的参与方,如何对新来的参与方进行公平合理的贡献评估.如何设计动态自适应的联邦学习公平性方法,保证联邦学习全生命周期的公平性,是未来研究的关键.

6.3 可信联邦学习不同维度之间的权衡

目前可信联邦学习大多是从公平性、隐私性、鲁棒或可解释性等单一的维度开展研究.然而,不同维度之间存在着某种程度上的一致性,也存在着互相冲突的情况.因此,在研究可信联邦学习时应关注不同维度之间的复杂交互作用.知道两个维度之间的一致性为我们带来了实现一个维度的替代思路:我们可以通过实现另一个维度来满足一个维度.此外,当两个维度相互矛盾时,我们可以根据需求在它们之间进行权衡.下面提出几个未来值得研究探索的结合点.

(1)公平性和隐私性. 隐私保护是联邦学习的主要目标,而公平性的研究可能会增加潜在的数据隐私安全隐患. 目前大部分方案都假设数据安全问题可以依托相关技术解决,但存在一些方法在设计角度对数据隐私安全缺乏考虑,甚至需移动参与方数据. 另外,存在恶意攻击者利用推理攻击等方式通过中间结果获得参与方的原始数据[129,130]. 因此,未来在方案设计中,需要研究者更多去了解联邦数据隐私安全相关技术,例如使用差分隐私技术,对计算结果施加扰动,防止隐私推理攻击. 此外,安全多方计算也可被用以保护客户端传输数据的安全,避免在方案中引入难以解决的数据隐私保护弊端.

(2)公平性和鲁棒性. 鲁棒性和公平性之间可能会互相冲突,无论是平均分配理论中的设计模型优化算法或修改全局聚合规则,还是按劳分配理论中的贡献评估和计算回报机制,都不可避免地会对其中一些客户端的性能造成影响,从而影响联邦学习整体的精度或效率. 因此,如何公平且高效地设计联邦学习算法变得尤为重要.

(3)公平性和可解释性. 在联邦学习中,可解释性和公平性的研究可能存在相互促进的关系. 从联邦学习客户的角度来看,他们缺乏确定是否被公平对待的机制. 这种不确定性可能会妨碍某些客户未来决定加入联邦学习. 构建可解释性的目标是提供全局理解,了解联邦学习服务器如何做出决策以及这些决策如何影响每个客户的利益,以建立双方之间的信任. 因此,通过研究可解释性能够细粒度地进行公平方案的设计. 另外,联邦学习公平性的研究需要对每个客户端有很多的先验知识. 无论是平均分配理论还是按劳分配理论都关注模型最终的性能. 然而,现有的方法还不能精确地控制并量化分配给每个联邦学习客户端模型的最终性能. 可解释性的研究可能会帮助解决这个问题.

虽然有许多问题需要研究,但理解不同维度之间的相互作用对于可信联邦学习非常重要. 如何集成公平性、隐私性、鲁棒性和可解释性等可信联邦学习不同维度,研究真正的可信联邦学习一体化机制和算法是未来研究的核心.

7　结语

随着数据流通和协同计算的不断发展,研究可信联邦学习中的公平性是大势所趋,提升联邦学习的公平性能够保证客户端参与的积极性和训练的可持续性. 目前,可信联邦学习公平性的研究仍然处于起步阶段,面临着许多挑战需要解决. 本文全面综述了可信联邦学习公平性的相关研究:首先,归纳梳理了联邦学习中存在的偏见以及不同应用场景中的公平性定义,并将其分为平均分配和按劳分配的联邦学习公平性两类;其次,根据不同的应用场景描述了各自的内涵和研究挑战,并详细探讨了联邦学习的公平性解决方案;然后,详细梳理了联邦学习公平性研究中常用的数据集、实验场景设置和公平评价指标;最后,展望了联邦学习公平性未来研究方向.

参考文献

[1] YANG Q, LIU Y, CHENG Y, et al. Federated learning[J]. Synthesis Lectures on Artificial in Telligence and Machine Learning, 2019, 13(3): 1-207.

[2] JAKUB KONEN, MCMAHAN H B, YU F X, et al. Learning pederated: Strategies for improving communication efficiency[EB/OL]. (2017-10-30)[2023-02-17]. https://arxiv.org/abs/1610.05492.

[3] DWORK C, HARDT M, PITASSI T, et al. Fairness through awareness[C]//Proceedings of the 3rd Innovations in Theoretical Computer Science Conference. Cambridge: ACM, 2012: 214-226.

[4] KAHNEMAN D, KNETSCH J L, THALER R H. Fairness and the assumptions of economics[J]. Journal of business, 1986, 59(4): S285-S300.

[5] FEHR E, SCHMIDT K M. A theory of fairness, competition, and cooperation[J]. The Quarterly Journal of Economics, 1999, 114(3): 817-868.

[6] JAIN R K, CHIU D M W, HAWE W R. A quantitative measure of fairness and discrimination For resource allocation in shared computer systems[EB/OL]. (1998-09-24)[2023-02-17]. https://arxiv.org/abs/cs/9809099.

[7] GABBAY D, PNUELI A, SHELAH S, et al. On the temporal analysis of fairness[C]//Proceedings of the 7th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages. Las Vegas: ACM, 1980: 163-173.

[8] MEHRABI N, MORSTATTER F, SAXENA N, et al. A survey on bias and fairness in machine learning[J]. ACM Computing Surveys (CSUR), 2021, 54(6): 1-35.

[9] CORBETT-DAVIES S, GOEL S. The measure and mismeasure of fairness: A critical review of fair machine learning[EB/OL]. [2023-02-17]. http://arxiv.org/abs/1808.00023.

[10] DU M, YANG F, ZOU N, et al. Fairness in deep learning: A computational perspective[J]. IEEE Intelligent Systems, 2020, 36(4): 25-34.

[11] GAJANE P, PECHENIZKIY M. On formalizing fairness in prediction with machine learning[EB/OL]. (2018-05-28)[2023-02-17]. https://arxiv.org/abs/1710.03184.

[12] RAJKOMAR A, HARDT M, HOWELL M D, et al. Ensuring fairness in machine learning to advance health eq-

uity[J]. Annals of Internal Medicine, 2018, 169(12): 866-872.

[13] 刘文炎, 沈楚云, 王祥丰, 等. 可信机器学习的公平性综述[J]. 软件学报, 2021, 32(5): 1404-1426.
LIU W Y, SHEN C Y, WANG X F, et al. Fairness in trustworthy machine learning: a survey[J]. Journal of Software, 2021, 32(5): 1404−1426. (in Chinese)

[14] MCMAHAN B, MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Artificial Intelligence and Statistics. Seattle: PMLR, 2017: 1273-1282.

[15] 张洪磊, 李浥东, 邬俊, 等. 基于隐私保护的联邦推荐算法综述[J]. 自动化学报, 2022, 48(9): 2142-2163.
ZHANG H L, LI Y D, WU J, et al. A survey on privacy-preserving federated recommender systems[J]. Acta Automatica Sinica, 2022, 48(9): 2142-2163. (in Chinese)

[16] DENG Y, LYU F, REN J, et al. Auction: Automated and quality-aware client selection framework for efficient federated learning[J]. IEEE Transactions on Parallel and Distributed Systems, 2021, 33(8): 1996-2009.

[17] UR REHMAN M H, DIRIR A M, SALAH K, et al. TrustFed: A framework for fair and trustworthy cross-device federated learning in IIoT[J]. IEEE Transactions on Industrial Informatics, 2021, 17(12): 8485-8494.

[18] HUANG W, LI T, WANG D, et al. Fairness and accuracy in horizontal federated learning[J]. Information Sciences, 2022, 589: 170-185.

[19] LI T, HU S, BEIRAMI A, et al. Ditto: Fair and robust federated learning through personalization[C]//International Conference on Machine Learning. Virtual Event: PMLR, 2021: 6357-6368.

[20] LYU L, XU X, WANG Q, et al. Collaborative fairness in federated learning[J]. Federated Learning: Privacy and Incentive, 2020, 12500: 189-204.

[21] MOHRI M, SIVEK G, SURESH A T. Agnostic federated learning[C]//International Conference on Machine Learning. Long Beach: PMLR, 2019: 4615-4625.

[22] LI T, SAHU A K, TALWALKAR A, et al. Federated learning: Challenges, methods, and future directions[J]. IEEE Signal Processing Magazine, 2020, 37(3): 50-60.

[23] ZHANG F, KUANG K, LIU Y, et al. Unified group fairness on federated learning[EB/OL]. (2022-02-16) [2023-02-17]. https://arxiv.org/abs/2111.04986.

[24] PAPADAKI A, MARTINEZ N, BERTRAN M, et al. Minimax demographic group fairness in federated learning[C]//2022 ACM Conference on Fairness, Accountability, and Transparency. Seoul: ACM, 2022: 142-159.

[25] FENG J, LIU L, PEI Q, et al. Min-max cost optimization for efficient hierarchical federated learning in wireless edge networks[J]. IEEE Transactions on Parallel and Distributed Systems, 2021, 33(11): 2687-2700.

[26] ZHU Z, SI S, WANG J, et al. Cali3F: Calibrated fast fair federated recommendation system[C]//2022 International Joint Conference on Neural Networks (IJCNN). Padua: IEEE, 2022: 1-8.

[27] HUANG W, LI T, WANG D, et al. Fairness and accuracy in federated learning[EB/OL]. (2020-12-18)[2023-02-17]. https://arxiv.org/abs/2012.10069.

[28] CUI S, PAN W, LIANG J, et al. Addressing algorithmic disparity and performance inconsistency in federated learning[J]. Advances in Neural Information Processing Systems, 2021, 34: 26091-26102.

[29] HORVATH S, LASKARIDIS S, ALMEIDA M, et al. Fjord: Fair and accurate federated learning under heterogeneous targets with ordered dropout[J]. Advances in Neural Information Processing Systems, 2021, 34: 12876-12889.

[30] NAGALAPATTI L, NARAYANAM R. Game of gradients: Mitigating irrelevant clients in federated learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Virtual Event: AAAI Press, 2021, 35(10): 9046-9054.

[31] WANG G, DANG C X, ZHOU Z. Measure contribution of participants in federated learning[C]//2019 IEEE International Conference on Big Data (Big Data). Los Angele: IEEE, 2019: 2597-2604.

[32] TANG S, GHORBANI A, YAMASHITA R, et al. Data valuation for medical imaging using Shapley value and application to a large-scale chest X-ray dataset[J]. Scientific reports, 2021, 11(1): 1-9.

[33] YOON J, ARIK S, PFISTER T. Data valuation using reinforcement learning[C]//International Conference on Machine Learning. Virtual Event: PMLR, 2020: 10842-10851.

[34] ZHANG J, WU Y, PAN R. Incentive mechanism for horizontal federated learning based on reputation and reverse auction[C]//Proceedings of the Web Conference. Virtual Event: ACM, 2021: 947-956.

[35] SHI Z, ZHANG L, YAO Z, et al. FedFAIM: A model performance-based fair incentive mechanism for federated learning[EB/OL]. (2022-07-16) [2023-02-17]. https://ieeexplore.ieee.org/document/9797864.

[36] SHI Y, LIU Z, SHI Z, et al. Fairness-aware client selection for federated learning[C]//2023 IEEE International Conference on Multimedia and Expo (ICME). Brisbane: IEEE, 2023: 324-329.

[37] ZHU H, ZHOU Y, QIAN H, et al. Online client selection for asynchronous federated learning with fairness consideration[J]. IEEE Transactions on Wireless Communications, 2022, 22(4): 2493-2506.

[38] NISHIO T, YONETANI R. Client selection for federated learning with heterogeneous resources in mobile edge[C]// ICC 2019-2019 IEEE International Conference on Communications (ICC). Shanghai: IEEE, 2019: 1-7.

[39] RIBERO M, VIKALO H. Communication-efficient federated learning via optimal client sampling[EB/OL]. (2020-10-14)[2023-02-17]. https://arxiv.org/abs/2007.15197.

[40] WANG H, KAPLAN Z, NIU D, et al. Optimizing federated learning on non-iid data with reinforcement learning [C]//IEEE INFOCOM 2020-IEEE Conference on Computer Communications. Toronto: IEEE, 2020: 1698-1707.

[41] YANG M, WANG X, ZHU H, et al. Federated learning with class imbalance reduction[C]//2021 29th European Signal Processing Conference (EUSIPCO). Dublin: IEEE, 2021: 2174-2178.

[42] LI C, ZENG X, ZHANG M, et al. PyramidFL: A fine-grained client selection framework for efficient federated learning[C]//Proceedings of the 28th Annual International Conference on Mobile Computing And Networking. Sydney: ACM, 2022: 158-171.

[43] HUANG T, LIN W, WU W, et al. An efficiency-boosting client selection scheme for federated learning with fairness guarantee[J]. IEEE Transactions on Parallel and Distributed Systems, 2020, 32(7): 1552-1564.

[44] HUANG T, LIN W, SHEN L, et al. Stochastic client selection for federated learning with volatile clients[J]. IEEE Internet of Things Journal, 2022, 9(20): 20055-20070.

[45] SULTANA A, HAQUE M M, CHEN L, et al. Eiffel: Efficient and fair scheduling in adaptive federated learning [J]. IEEE Transactions on Parallel and Distributed Systems, 2022, 33(12): 4282-4294.

[46] WANG H, QU Z, GUO S, et al. Intermittent pulling with local compensation for communication-efficient distributed learning[J]. IEEE Transactions on Emerging Topics in Computing, 2020, 10(2): 779-791.

[47] HAO W, EL-KHAMY M, LEE J, et al. Towards fair federated learning with zero-shot data augmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Virtual Event: IEEE, 2021: 3310-3319.

[48] LI T, SANJABI M, BEIRAMI A, et al. Fair resource allocation in federated learning[EB/OL]. (2020-12-23)[2023-02-17]. http://arxiv.org/abs/1905.10497.

[49] DU W, XU D, WU X, et al. Fairness-aware agnostic federated learning[C]//Proceedings of the 2021 SIAM International Conference on Data Mining (SDM). Virtual Event: SIAM, 2021: 181-189.

[50] LI T, SAHU A K, ZAHEER M, et al. Federated optimization in heterogeneous networks[J]. Proceedings of Machine Learning and Systems, 2020, 2: 429-450.

[51] HU Z, SHALOUDEGI K, ZHANG G, et al. Federated learning meets multi-objective optimization[J]. IEEE Transactions on Network Science and Engineering, 2022, 9(4): 2039-2051.

[52] KARIMIREDDY S P, KALE S, MOHRI M, et al. Scaffold: Stochastic controlled averaging for federated learning[C]//International Conference on Machine Learning. Virtual Event: PMLR, 2020: 5132-5143.

[53] YUROCHKIN M, AGARWAL M, GHOSH S, et al. Bayesian nonparametric federated learning of neural networks[C]//International Conference on Machine Learning. Long Beach: PMLR, 2019: 7252-7261.

[54] WANG H, YUROCHKIN M, SUN Y, et al. Federated learning with matched averaging[EB/OL]. (2020-05-02)[2023-02-17]. https://arxiv.org/abs/2002.06440.

[55] ZENG H, ZHOU T, GUO Y, et al. FedCav: Contribution-aware model aggregation on distributed heterogeneous data in federated learning[C]//50th International Conference on Parallel Processing. Lemont: ACM, 2021: 1-10.

[56] WANG Z, FAN X, QI J, et al. Federated learning with fair averaging[C]//Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence. Virtual Event: ijcai.org, 2021: 1615-1623.

[57] FRABONI Y Y, VIDAL R, LORENZI M. Free-rider attacks on model aggregation in federated learning[C]//International Conference on Artificial Intelligence and Statistics. Virtual Event: PMLR, 2021: 1846-1854.

[58] FAN Z, FANG H, ZHOU Z, et al. Improving fairness for data valuation in horizontal federated learning[C]//2022 IEEE 38th International Conference on Data Engineering (ICDE). Kuala Lumpur: IEEE, 2022: 2440-2453.

[59] YANG C, LIU J, SUN H, et al. WTDP-Shapley: Efficient and effective incentive mechanism in federated learning for intelligent safety inspection[EB/OL]. (2022-08-16)[2023-02-17]. https://ieeexplore. ieee. org/abstract/document/9857582.

[60] WEI S, TONG Y, ZHOU Z, et al. Efficient and fair data valuation for horizontal federated learning[J]. Federated Learning: Privacy and Incentive, 2020, 12500: 139-152.

[61] FAN Z, FANG H, ZHOU Z, et al. Fair and efficient contribution valuation for vertical federated learning [EB/OL]. (2022-01-07)[2023-02-17]. https://arxiv.org/abs/2201.02658.

[62] WINTER E. The shapley value[J]. Handbook of Game Theory with Economic Applications, 2002, 3: 2025-2054.

[63] LITTLECHILD S C, OWEN G. A simple expression for the Shapley value in a special case[J]. Management Science, 1973, 20(3): 370-372.

[64] KALAI E, SAMET D. On weighted Shapley values[J]. In-

ternational Journal of Game Theory, 1987, 16(3): 205-222.
[65] MERRICK L, TALY A. The explanation game: Explaining machine learning models using shapley values[C]// Machine Learning and Knowledge Extraction: 4th IFIP TC 5, TC 12, WG 8.4, WG 8.9, WG 12.9 International Cross-Domain Conference, CD-MAKE 2020. Dublin: Springer, 2020: 17-38.
[66] SUNDARARAJAN M, NAJMI A. The many Shapley values for model explanation[C]//International Conference on Machine Learning. Virtual Event: PMLR, 2020: 9269-9278.
[67] KUMAR I E, VENKATASUBRAMANIAN S, SCHEIDEGGER C, et al. Problems with Shapley-value-based explanations as feature importance measures[C]//International Conference on Machine Learning. Virtual Event: PMLR, 2020: 5491-5500.
[68] ROZEMBERCZKI B, WATSON L, BAYER P, et al. The shapley value in machine learning[C]//Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence. Vienna: ijcai.org, 2022: 5572-5579.
[69] WANG G. Interpret federated learning with shapley values[EB/OL]. (2019-05-28) [2023-02-17]. http://arxiv. org/abs/1905.04519.
[70] HUANG J, TALBI R, ZHAO Z, et al. An exploratory analysis on users' contributions in federated learning[C]// 2020 Second IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA). Atlanta: IEEE, 2020: 20-29.
[71] LIM W Y B, HUANG J, XIONG Z, et al. Towards federated learning in uav-enabled internet of vehicles: A multi-dimensional contract-matching approach[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22 (8): 5140-5154.
[72] HUANG J, HONG C, CHEN L Y, et al. Is shapley value fair? Improving client selection for mavericks in federated learning[EB/OL]. (2021-07-29) [2023-02-17]. https://arxiv.org/abs/2106.10734.
[73] GHORBANI A, ZOU J. Data shapley: Equitable valuation of data for machine learning[C]//International Conference on Machine Learning. Long Beach: PMLR, 2019: 2242-2251.
[74] JIA R, DAO D, WANG B, et al. Towards efficient data valuation based on the shapley value[C]//The 22nd International Conference on Artificial Intelligence and Statistics. Long Beach: PMLR, 2019: 1167-1176.
[75] LIU Z, CHEN Y, YU H, et al. Gtg-shapley: Efficient and accurate participant contribution evaluation in federated learning[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2022, 13(4): 1-21.
[76] WANG T, RAUSCH J, ZHANG C, et al. A principled approach to data valuation for federated learning[J]. Federated Learning: Privacy and Incentive, 2020, 12500: 153-167.
[77] XU X, LYU L, MA X, et al. Gradient driven rewards to guarantee fairness in collaborative machine learning[J]. Advances in Neural Information Processing Systems, 2021, 34: 16104-16117.
[78] ZHANG J, LI C, ROBLES-KELLY A, et al. Hierarchically fair federated learning[EB/OL]. (2020-04-28)[2023-02-17]. https://arxiv.org/abs/2004.10386.
[79] CHEN Y, YANG X, QIN X, et al. Dealing with label quality disparity in federated learning[J]. Federated Learning: Privacy and Incentive, 2020, 12500: 108-121.
[80] XU X, LYU L. A reputation mechanism is all you need: Collaborative fairness and adversarial robustness in federated learning[EB/OL]. (2021-07-27)[2023-02-17]. https://arxiv.org/abs/2011.10464.
[81] LYU L, YU J, NANDAKUMAR K, et al. Towards fair and privacy-preserving federated deep models[J]. IEEE Transactions on Parallel and Distributed Systems, 2020, 31(11): 2524-2541.
[82] PANDEY S R, TRAN N H, BENNIS M, et al. A crowdsourcing framework for on-device federated learning[J]. IEEE Transactions on Wireless Communications, 2020, 19(5): 3241-3256.
[83] YAN B, LIU B, WANG L, et al. Fedcm: A real-time contribution measurement method for participants in federated learning[C]//2021 International Joint Conference on Neural Networks (IJCNN). Shenzhen: IEEE, 2021: 1-8.
[84] NISHIO T, SHINKUMA R, MANDAYAM N B. Estimation of individual device contributions for incentivizing federated learning[C]//2020 IEEE Globecom Workshops (GC Wkshps). Virtual Event: IEEE, 2020: 1-6.
[85] ZHAO J, ZHU X, WANG J, et al. Efficient client contribution evaluation for horizontal federated learning[C]// ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Toronto: IEEE, 2021: 3060-3064.
[86] ZHAN Y, ZHANG J, HONG Z, et al. A survey of incentive mechanism design for federated learning[J]. IEEE Transactions on Emerging Topics in Computing, 2021, 10 (2): 1035-1044.
[87] ZHAN Y, ZHANG J, LI P, et al. Crowdtraining: Architecture and incentive mechanism for deep learning training in the internet of things[J]. IEEE Network, 2019, 33(5): 89-95.
[88] ZHAN Y, LI P, QU Z, et al. A learning-based incentive mechanism for federated learning[J]. IEEE Internet of Things Journal, 2020, 7(7): 6360-6368.
[89] CONG M, YU H, WENG X, et al. A game-theoretic framework for incentive mechanism design in federated

learning[J]. Federated Learning: Privacy and Incentive, 2020, 12500: 205-222.

[90] TU X, ZHU K, LUONG N C, et al. Incentive mechanisms for federated learning: From economic and game theoretic perspective[J]. IEEE Transactions on Cognitive Communications and Networking, 2022, 8(3): 1566-1593.

[91] DING N, FANG Z, HUANG J. Optimal contract design for efficient federated learning with multi-dimensional private information[J]. IEEE Journal on Selected Areas in Communications, 2020, 39(1): 186-200.

[92] UR REHMAN M H, SALAH K, DAMIANI E, et al. Towards blockchain-based reputation-aware federated learning[C]//IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). Toronto: IEEE, 2020: 183-188.

[93] TOYODA K, ZHANG A N. Mechanism design for an incentive-aware blockchain-enabled federated learning platform[C]//2019 IEEE international conference on big data (Big Data). Los Angeles: IEEE, 2019: 395-403.

[94] BAO X, SU C, XIONG Y, et al. Flchain: A blockchain for auditable federated learning with trust and incentive [C]//2019 5th International Conference on Big Data Computing and Communications (BIGCOM). Qingdao: IEEE, 2019: 151-159.

[95] SONG T, TONG Y, WEI S. Profit allocation for federated learning[C]//2019 IEEE International Conference on Big Data (Big Data). Los Angeles: IEEE, 2019: 2577-2586.

[96] SIM R H L, ZHANG Y, CHAN M C, et al. Collaborative machine learning with incentive-aware model rewards[C] //International Conference on Machine Learning. Virtual Event: PMLR, 2020: 8927-8936.

[97] YU H, LIU Z, LIU Y, et al. A fairness-aware incentive scheme for federated learning[C]//Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society. New York: ACM, 2020: 393-399.

[98] YU H, LIU Z, LIU Y, et al. A sustainable incentive scheme for federated learning[J]. IEEE Intelligent Systems, 2020, 35(4): 58-69.

[99] LIU Y, TIAN M, CHEN Y, et al. A contract theory based incentive mechanism for federated learning[M]//Federated and Transfer Learning. Cham: Springer International Publishing, 2022: 117-137.

[100] ZENG R, ZHANG S, WANG J, et al. Fmore: An incentive scheme of multi-dimensional auction for federated learning in mec[C]//2020 IEEE 40th International Conference on Distributed Computing Systems (ICDCS). Singapore: IEEE, 2020: 278-288.

[101] 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.

[102] LIM W Y B, XIONG Z, MIAO C, et al. Hierarchical incentive mechanism design for federated machine learning in mobile networks[J]. IEEE Internet of Things Journal, 2020, 7(10): 9575-9588.

[103] SARIKAYA Y, ERCETIN O. Motivating workers in federated learning: A stackelberg game perspective[J]. IEEE Networking Letters, 2019, 2(1): 23-27.

[104] KHAN L U, PANDEY S R, TRAN N H, et al. Federated learning for edge networks: Resource optimization and incentive mechanism[J]. IEEE Communications Magazine, 2020, 58(10): 88-93.

[105] ZHAN Y, ZHANG J. An incentive mechanism design for efficient edge learning by deep reinforcement learning approach[C]//IEEE INFOCOM 2020-IEEE Conference on Computer Communications. Toronto: IEEE, 2020: 2489-2498.

[106] ZHANG P, WANG C, JIANG C, et al. Deep reinforcement learning assisted federated learning algorithm for data management of IIoT[J]. IEEE Transactions on Industrial Informatics, 2021, 17(12): 8475-8484.

[107] CHEN S, SHEN C, ZHANG L, et al. Dynamic aggregation for heterogeneous quantization in federated learning [J]. IEEE Transactions on Wireless Communications, 2021, 20(10): 6804-6819.

[108] LE T H T, TRAN N H, TUN Y K, et al. An incentive mechanism for federated learning in wireless cellular networks: An auction approach[J]. IEEE Transactions on Wireless Communications, 2021, 20(8): 4874-4887.

[109] LO S K, LIU Y, LU Q, et al. Blockchain-based trustworthy federated learning architecture[J]. IEEE Internet of Things Journal, 2022, 10(4): 3276-3284.

[110] WANG Z, HU Q, LI R, et al. Incentive mechanism design for joint resource allocation in blockchain-based federated learning[J]. IEEE Transactions on Parallel and Distributed Systems, 2023, 34(5): 1536-1547.

[111] CHAI H, LENG S, CHEN Y, et al. A hierarchical blockchain-enabled federated learning algorithm for knowledge sharing in internet of vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 22(7): 3975-3986.

[112] PENG Z, XU J, CHU X, et al. Vfchain: Enabling verifiable and auditable federated learning via blockchain systems[J]. IEEE Transactions on Network Science and Engineering, 2021, 9(1): 173-186.

[113] ZHANG C, ZHU L, XU C, et al. BSFP: Blockchain-enabled smart parking with fairness, reliability and privacy protection[J]. IEEE Transactions on Vehicular Technolo-

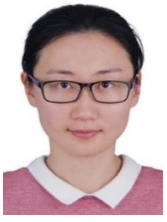
gy, 2020, 69(6): 6578-6591.
[114] KIM H, PARK J, BENNIS M, et al. Blockchained on-device federated learning[J]. IEEE Communications Letters, 2019, 24(6): 1279-1283.
[115] WENG J, WENG J, ZHANG J, et al. Deepchain: Auditable and privacy-preserving deep learning with blockchain-based incentive[J]. IEEE Transactions on Dependable and Secure Computing, 2019, 18(5): 2438-2455.
[116] 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, 2020, 8(7): 5926-5937.
[117] HAN T, GONG X, FENG F, et al. Privacy-preserving multi-source domain adaptation for medical data[J]. IEEE Journal of Biomedical and Health Informatics, 2022, 27(2): 842-853.
[118] KALRA S, WEN J, CRESSWELL J C, et al. Decentralized federated learning through proxy model sharing[J]. Nature Communications, 2023, 14(1): 2899.
[119] CALDAS S, DUDDU S M K, WU P, et al. Leaf: A benchmark for federated settings[EB/OL]. (2020-12-23) [2023-02-17]. http://arxiv.org/abs/1812.01097.
[120] COHEN G, AFSHAR S, TAPSON J, et al. EMNIST: Extending MNIST to handwritten letters[C]//2017 International Joint Conference on Neural Networks (IJCNN). Anchorage: IEEE, 2017: 2921-2926.
[121] XIAO H, RASUL K, VOLLGRAF R. Fashion-mnist: A novel image dataset for benchmarking machine learning algorithms[EB/OL]. (2018-08-13) [2023-02-17]. http://arxiv.org/abs/1708.07747.
[122] DARLOW L N, CROWLEY E J, ANTONIOU A, et al. Cinic-10 is not imagenet or cifar-10[EB/OL]. (2018-10-30)[2023-02-17]. http://arxiv.org/abs/1810.03505.
[123] LI Q, DIAO Y, CHEN Q, et al. Federated learning on non-iid data silos: An experimental study[C]//2022 IEEE 38th International Conference on Data Engineering (ICDE). Kuala Lumpur: IEEE, 2022: 965-978.
[124] KAIROUZ P, MCMAHAN H B, AVENT B, et al. Advances and open problems in federated learning[J]. Foundations and Trends in Machine Learning, 2021, 14(1-2): 1-210.
[125] ZHAO Z, FENG C, HONG W, et al. Federated learning with non-iid data in wireless networks[J]. IEEE Transactions on Wireless communications, 2021, 21(3): 1927-1942.
[126] TAN A Z, YU H, CUI L, et al. Towards personalized federated learning[EB/OL]. (2022-07-19) [2023-02-17]. https://arxiv.org/abs/2103.00710.
[127] JAMALI-RAD H, ABDIZADEH M, SINGH A. Federated learning with taskonomy for non-IID data[EB/OL]. (2021-04-07)[2023-02-17]. https://arxiv.org/abs/2103.15947.
[128] ZHAO Z, JOSHI G. A dynamic reweighting strategy for fair federated learning[C]//ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Virtual Event, Singapore: IEEE, 2022: 8772-8776.
[129] CAO X, FANG M, LIU J, et al. FLTrust: Byzantine-robust federated learning via trust bootstrapping[C]//Proceeding 2021 Network and Distributed System Security Symposium (NDSS). Reston: Internet Society, 2021: 24434.
[130] FANG M, CAO X, JIA J, et al. Local model poisoning attacks to byzantine-robust federated learning[C]//Proceedings of the 29th USENIX Conference on Security Symposium. Virtual Event: USENIX Association, 2020: 1623-1640.

作者简介

陈颢瑜 男,1995年1月出生于河南省洛阳市. 北京交通大学计算机与信息技术学院博士研究生. 主要研究方向为隐私计算与可信智能.
E-mail: hychen95@bjtu.edu.cn

李浥东(通讯作者) 男,1982年2月出生于山西省太原市. 北京交通大学计算机与信息技术学院教授. 主要研究方向为大数据分析与安全、数据隐私保护与先进计算.
E-mail: ydli@bjtu.edu.cn

张洪磊 男,1993年7月出生于河北省邢台市. 北京交通大学计算机与信息技术学院博士研究生. 主要研究方向为推荐系统与隐私保护.
E-mail: honglei.zhang@bjtu.edu.cn



陈乃月 女,1989年8月出生于河北省唐山市. 北京交通大学计算机与信息技术学院讲师. 主要研究方向为社交网络、数据挖掘与联邦学习.
E-mail: nychen@bjtu.edu.cn