软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software, 2023,34(3):1168−1192 [doi: 10.13328/j.cnki.jos.006786]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

联邦学习贡献评估综述*

王 勇, 李国良, 李开宇

(清华大学 计算机科学与技术系, 北京 100084)

通信作者: 李国良, E-mail: liguoliang@tsinghua.edu.cn

摘 要: 数据不动的联邦学习框架是多个数据持有方合作训练机器学习模型的新范式. 多个数据持有方参与联邦学习时的贡献评估是联邦学习的核心问题之一. 参与方贡献评估需要兼顾有效性、公平性和合理性等要素, 在理论方法与实际应用中均面临多项挑战. 贡献评估首先需要明确如何度量数据价值, 然而数据估值存在主观性与依赖于实际任务场景的特点, 如何设计有效、可靠并对恶意数据鲁棒的数据估值指标是第一大挑战. 其次, 联邦学习合作中的参与方贡献评估是经典的合作博弈问题, 如何制定公平合理的参与方贡献评估方案, 实现参与方一致认可的博弈平衡是第二大挑战. 最后, 参与方贡献评估往往计算复杂度高, 同时, 联邦学习中围绕模型的数据估值时间开销大, 因此, 在实践中如何设计高效且准确的近似算法是第三大挑战. 近年来, 为了有效地解决上述挑战, 学术界对联邦学习中的贡献评估问题展开了广泛的研究. 首先, 简要介绍联邦学习与参与方贡献评估的背景知识; 然后, 综述数据估值指标、参与方贡献评估方案和相关优化技术; 最后, 讨论了联邦学习贡献评估仍面临的挑战并展望未来研究的发展方向.

关键词: 贡献评估; 数据估值; 联邦学习; 激励机制; 合作博弈

中图法分类号: TP18

中文引用格式: 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

英文引用格式: 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). http://www.jos.org.cn/1000-9825/6786.htm

Survey on Contribution Evaluation for Federated Learning

WANG Yong, LI Guo-Liang, LI Kai-Yu

(Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

Abstract: Federated learning is a collaborative machine learning framework with multiple participants whose training datasets are kept locally. How to evaluate the corresponding data contribution of each participant is one of the critical problems of federated learning. However, contribution evaluation in federated learning faces multiple challenges. First, to evaluate participant contribution, data value needs to be quantified, however, data valuation is challenging because it is subjective, task context-dependent, and vulnerable to malicious data. Second, participant contribution evaluation is a classic cooperative game problem, and a fair yet rational cooperative contribution evaluation scheme is needed to achieve an optimal equilibrium among all participants. Third, contribution evaluation schemes often involve exponential computational complexity, where data valuation by training models in federated learning is also quite time consuming. In recent years, researchers have conducted extensive studies on participant contribution evaluation in federated learning to tackle the above challenges. This study first introduces the background knowledge of federated learning and contribution evaluation. Then, data valuation metrics, contribution evaluation schemes, and corresponding optimization technologies are surveyed successively. Finally, the remaining challenges of contribution evaluation and potential future work are discussed.

Key words: contribution evaluation; data valuation; federated learning; incentive mechanism; cooperative game

* 基金项目: 国家自然科学基金(61925205); 北京国家信息研究中心资助项目
本文由“大数据治理的理论与技术”专题特约编辑杜小勇教授、杨晓春教授和童咏昕教授推荐.
收稿时间: 2022-05-15; 修改时间: 2022-07-29, 2022-09-07; 采用时间: 2022-09-23; jos 在线出版时间: 2022-10-27

随着人工智能相关技术的广泛应用, 驱动人工智能模型的大数据已成为信息时代的“石油”[1]. 然而, 数据孤岛的存在, 阻碍了大数据赋能应用. 从粗粒度来看, 数据被掌握在多个服务提供商手中; 从细粒度来看, 数据产生于成千上万个个人或者物联网设备端侧. 为了解决数据孤岛下的数据赋能问题, 近年来, 多个数据持有方合作训练机器学习模型的方式逐渐流行起来[2]. 但是, 由多个数据持有方提供数据、基于中心服务器融合多方数据来训练模型的方案存在数据隐私安全和参与方权益保护等多方面的问题. 为了应对相关挑战, “数据不动模型动”的联邦学习框架随之兴起[3], 并逐渐成为多方合作训练模型的新范式. 联邦学习中, 各参与方每轮次传递隐私安全的模型梯度更新来合作训练联邦模型, 各参与方数据不离开本地, 保证了数据隐私安全.

为了吸引潜在优质数据持有方参与联邦学习, 需要根据各参与方数据在联邦合作中的预计贡献大小给予激励. 为此, 如何有效地评估参与方(数据)在联邦合作中的贡献, 是联邦学习实际应用与长远发展的关键问题. 贡献评估首先需要考虑如何度量数据价值高低的问题, 即数据估值子问题. 最直观的数据估值指标是测试准确率: 基于数据训练联邦模型, 在联邦测试集上评测模型, 测试准确率越高, 则认为训练数据的价值越高. 确定数据估值指标后, 容易将参与方自身的数据价值直接等同于参与方对联邦合作的贡献, 比如, 用参与方独自训练模型的联邦测试准确率来评价其在后续联邦合作中的贡献. 然而, 联邦合作下全体参与方的数据价值一般不等于各参与方数据价值之和, 即参与方个体数据价值并不能代表其在联邦合作中的贡献. 从博弈论角度来看, 评估个体在联邦合作中的贡献是合作博弈问题[4], 需要考虑如何实现利益的均衡分配. 比如, 持有常规场景数据的参与方个体数据价值大, 而持有少量互补场景数据的参与方个体数据价值小, 但在已有常规参与方情况下, 互补参与方能为联邦带来显著的边际价值增益, 解决联邦性能瓶颈的问题. 因此, 贡献评估需要从联邦集体出发, 研究参与方数据在组合中的贡献期望, 以有效地评价参与方在联邦合作中的贡献, 即制定贡献评估方案子问题. 此外, 贡献评估需要针对不同数据进行多次价值度量, 而联邦学习价值度量往往需要重新训练并评测模型, 存在效率低下等性能优化子问题. 解决上述问题主要面临如下 3 个方面的挑战.

- 数据估值的有效性和可靠性. 如何度量数据价值是贡献评估的首要问题, 然而数据估值存在主观性, 且严重依赖于任务上下文, 比如对异常监测任务高价值的数据, 可能对于回归或分类任务价值不高. 同时, 在实践中可能存在策略性参与方的情况, 这些参与方希望通过低质量或者恶意数据从联邦中获利, 比如通过复制数据期望提高获利. 因此, 数据估值指标需要对这些低价值或者恶意数据具备鲁棒性, 识别并降低其对估值带来的影响. 当任务场景明确且任务测试集完备时, 可通过联邦模型在测试集上的效果来度量数据价值. 然而, 实践中很难收集到完备的测试数据, 且任务测试效果受到任务难度、模型有效性等因素的影响[5], 因此很难设计有效可靠的数据估值指标.
- 贡献评估的公平性和合理性. 从联邦集体来看, 参与方个体的数据价值不能代表其在联邦合作中的贡献, 贡献评估需要从参与方组合的数据价值出发, 评估引入某个参与方为联邦带来的边际价值增益. 从参与方个体来看, 在参与方联邦合作中, 贡献评估需要考虑公平性的问题, 比如, 持有类似数据的参与方应有类似的贡献评估结果, 在任何其他参与方组合下引入某参与方均不能给联邦合作带来边际价值增益, 则该参与方对联邦无贡献. 同时, 从联邦合作组合合理性角度来看, 全体参与方合作中, 数据的组合价值应等同于所有参与方的贡献之和. 此外, 贡献评估需要考虑新增数据价值度量指标、参与方组合的贡献评估结果相对于其数据价值是否足够合理等问题.
- 贡献评估计算的高效性. 有效的贡献评估方案往往需要枚举不同参与方组合来度量价值, 以计算引入某个参与方对联邦模型性能的边际提升效果. 然而, 潜在参与方组合数随参与方数量指数性增长. 同时, 联邦学习中不同参与方组合的数据估值往往需要重复训练并评测联邦模型, 这也额外地增加了计算代价. 因此, 如何围绕联邦学习特性设计高效的贡献评估优化算法, 是方案实际落地的关键问题. 此外, 低价值或者恶意参与方会对联邦带来负面的影响, 使得联邦合作训练的模型可用性降低.

自联邦学习被提出并广泛研究以来, 学术界已经开始研究参与方贡献评估的相关技术来攻克上述挑战.

- 首先, 解决贡献评估问题需要设计有效且可靠的数据估值指标, 以准确地度量数据对联邦学习任务的价值. 本文按是否依赖于联邦测试集综述两类数据估值指标. 对于联邦持有任务完备测试集的情

况，数据估值指标可以是数据训练联邦模型的测试准确率；从信息论角度，数据为测试集所带来的信息增益，等等．在联邦不具备测试集时，数据估值指标可以是任务相关的数据统计指标、数据训练任务模型的相似度和模型不确定性等．

- 其次，贡献评估的核心问题是探索公平合理的贡献评估方案，以客观评价参与方在联邦合作中的贡献．本文首先综述简单直观的个体法和留一法方案，并分析这两种方案在评估参与方联邦合作贡献的不足之处．本文接下来综述在合作博弈理论上更加公平合理的夏普利值和最小核方案，对比其相对于个体法和留一法的优势，并总结贡献评估方案相关的重要性质．
- 最后，围绕一些贡献评估方案固有的高计算复杂度问题，本文综述贡献评估计算的近似优化技术．本文首先综述统计采样优化方法，对比不同采样方法的理论计算复杂度，并调研实际应用中加速采样计算收敛的优化策略．随后，本文综述面向联邦学习的估值计算优化方法，调研如何复用模型梯度避免重复训练模型，并从随机采样、模型训练和相关数据这 3 个层面总结模型训练的剪枝策略．此外，本文综述如何基于贡献评估结果来优化联邦学习流程，分析如何降低低价值或恶意参与方对联邦带来的影响，保障联邦模型的高可用性．

图 1 展示了联邦学习贡献评估的相关技术，并以夏普利值方案为例，展示了数据估值与参与方贡献评估的关系．夏普利值方案中，参与方的贡献是其在联邦合作中的边际价值增益的期望，即枚举所有不包含某个参与方的参与组合下，计算引入该参与方带来的组合数据价值边际增益的平均值(将在第 3 节详细介绍)．

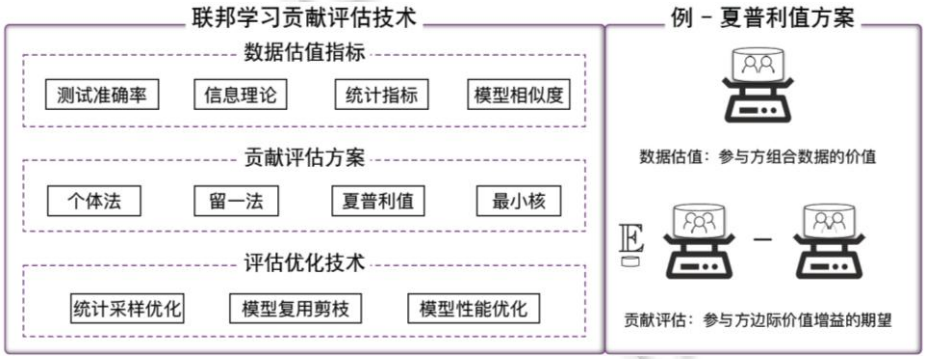

图 1　联邦学习贡献评估技术

本文主要综述针对联邦学习参与方的贡献评估技术，相关综述工作涵盖联邦学习参与方激励、数据定价、质量评估、隐私安全以及贡献评估方法性能实证等[6]．联邦学习参与方激励机制综述主要调研如何基于参与方数据贡献和算力、参与成本和平台利益最大化等因素，设计参与方在合作中的回报机制[7,8]．参与方对联邦的贡献是激励判定的关键要素之一，因此，贡献评估是联邦学习激励机制中的关键前置问题．考虑到实际平台构建的场景，数据定价综述从经济学原理和市场供需关系的角度调研了数据定价机制[9,10]，是联邦贡献转化为实际经济回报的重要参考．从数据管理的角度出发，数据质量评估综述调研质量评估与提升的方法，比如数据完整性、精度、独特性和时效性等数据固有特性[11]，并分析数据质量对机器学习任务的影响[12]．这些数据质量评价标准常被用于构造测试集无关的数据估值指标．从数据安全的角度出发，联邦学习数据隐私保护综述调研差分隐私、同态加密等技术以及对联邦学习可能的攻击形式[13]，这些技术对于本文中贡献评估方案的实际落地亦起到关键作用．此外，也有贡献评估相关的实证工作，验证对比了贡献评估技术的实际性能，比如在近似优化下，夏普利值方案在扩展性和可用性上可以同时超过留一法[14]；在牺牲一定效率时，基于贡献评估方案可以有效识别低贡献和恶意参与方等[15]．

本文第 1 节介绍联邦学习与贡献评估的背景知识．第 2 节介绍如何度量数据价值，并总结对比各类数据估值指标适应的场景．第3节介绍参与方贡献评估的方案，并总结分析各种博弈方案所满足的重要性质，对比不同方案优劣和适用场景．第 4 节调研面向联邦学习的贡献评估优化技术，从统计采样优化、模型复用剪枝

和降低恶意参与方对模型性能影响等方面详细梳理总结了相关技术. 第 5 节总结现有方法的不足之处, 展望未来数据估值指标设计的参考准则、贡献评估方案需考虑的要素以及面向联邦学习的评估优化研究方向.

1 背景知识

在综述调研联邦学习参与方贡献评估相关技术之前, 本节先介绍联邦学习以及贡献评估的背景知识.

1.1 联邦学习

联邦学习(federated learning, FL)是数据不动的分布式机器学习框架, 机器学习模型由多个参与方合作经过多轮次训练完成[16]. 如图 2 所示, 联邦学习由一个联邦服务器和多个参与方组成, 每个参与方持有各自的本地数据, 过程中, 各参与方的数据均不会离开参与方本地. 联邦模型经过多轮次的训练, 直至满足要求后终止, 比如测试准确率达到某个指标, 或者训练轮次超过限制等.

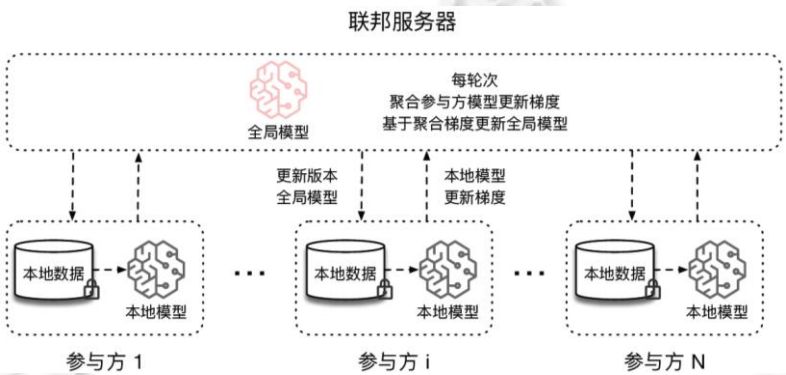

图 2 联邦学习框架

每轮次训练分为两个阶段.

(1) 本地更新: 各参与方基于各自的本地数据更新本地模型, 并将参数梯度更新发送给联邦服务器.

(2) 联邦聚合: 联邦服务器聚合所有参与方的更新. 比如, 联邦平均(FedAvg)简单基于各参与方数据量加权聚合各参与方上传的梯度值, 并基于聚合梯度更新全局模型.

最后, 联邦服务器将更新完成的全局模型参数发送给各参与方, 各参与方更新本地模型, 以准备下一轮次训练. 联邦学习也有去中心化的端到端架构, 由于本文调研的相关几乎全部基于中心服务器架构的设定[17], 因此, 本文对基于区块链技术的去中心化架构不做展开介绍[18,19]. 联邦学习定义一般包含如下假设[20].

- 多方参与: 有 2 到多个参与方合作训练联邦模型, 每个参与方及联邦平台根据利益需要作出合理的决策.
- 数据不动: 联邦参与方注重自身数据隐私与安全, 在合作训练联邦模型过程中, 任何参与方的原始数据不会以任何形式部分或全部地离开参与方本地.
- 可信传输: 联邦学习过程中, 梯度与参数等信息基于安全方式传递, 其中, 联邦中心服务器是诚实可信的(honest), 会按联邦学习要求完成训练并不会窃取参与方的隐私数据; 参与方是半诚实的(semi-honest, honest-but-curious), 会基于本地数据上传每轮次的真实梯度更新, 但对于获取其他参与方的隐私信息存在好奇[21].

按照参与方持有的数据维度不同, 联邦学习分为横向联邦学习(horizontal federated learning, HFL)和纵向联邦学习(vertical federated learning, VFL)[3].

- 横向联邦学习. 在横向联邦学习中, 各参与方拥有相同数据特征列下的不同数据行, 比如按照相同或者类似数据模型设计数据库的多个银行, 它们拥有相同的数据特征, 但是它们服务的客户不尽相同, 拥有不同的数据样本.

- 纵向联邦学习. 在纵向联邦学习中, 各参与方拥有相同数据行的不同数据列, 即持有不同数据特征, 比如为社会提供不同服务的银行与通信运营商, 它们拥有用户在银行信用与通信服务上的不同数据特征, 数据行可以通过用户身份信息对齐.

图 3 展示了横向与纵向联邦学习的差异. 其中, 基于参与方 A 和 B 相同数据特征列{年收入,贷款额}, 以及不同数据行[1001–1004]和[1003–1006]的场景为横向联邦学习; 基于参与方 A 和 B 的相同数据本行[1003–1004]以及不同数据特征列{性别,学历,年收入,贷款额}和{年收入,贷款额,岗位,居住地}的场景为纵向联邦学习. 横向与纵向联邦学习的分类设计较为理想, 然而在实际情况中, 不同参与方持有数据不能完全按照数据行或者列对齐的情况, 但是这种新兴的针对混合设定的迁移联邦学习研究尚未成熟, 本文调研的贡献评估方法未采用这种设定, 因此对迁移联邦学习不展开介绍[22].

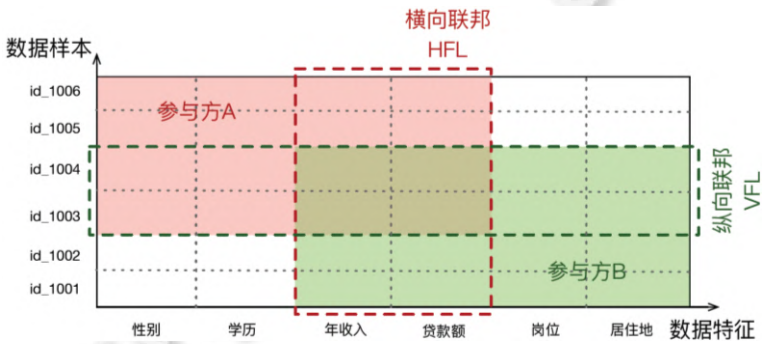

图 3 横向与纵向联邦学习对比

按照参与方对象的不同, 联邦学习被分为跨设备联邦学习(cross-device FL)和跨机构联邦学习(cross-silo FL)两类场景[17].

- 跨设备联邦学习. 个人终端或者物联网设备作为参与方的联邦学习称为跨设备联邦学习, 也称为企业对消费者联邦学习(business-to-consumer, B2C FL). 跨设备联邦学习场景呈现参与方数量多、算力有限、单个参与方数据量少、传输开销大等特点. 典型案例是谷歌和苹果等公司面向用户的智能手机应用训练联邦模型, 用户通过参与横向联邦学习来获得更智能的体验[23,24]. 然而, 即使在数据隐私保护下, 参与方仍然难有主观意愿参与联邦合作, 需要评估参与贡献, 对参与进行额外有效的激励.
- 跨机构联邦学习. 企业或结构作为参与方的联邦学习称为跨机构联邦学习, 也称为企业对企业联邦学习(business-to-business, B2B FL). 跨机构联邦学习呈现参与方数量少、算力充足、单个参与方数据量大、存在商业风险顾虑等特点. 典型案例是 FATE 联邦学习平台, 为金融等行业提供数据隐私保护的联邦学习方案[3]. 在跨机构联邦学习中, 参与方一般期望通过联邦合作获得性能更高的任务模型或者根据贡献获得金钱等回报. 为了避免低贡献参与方窃取联邦合作成果, 以及恶意参与方通过复制数据等策略以期望提升回报, 需要研究公平、有效且免疫低贡献或恶意参与方攻击的贡献评估方案.

1.2 参与方贡献评估

为了激励参与方加入联邦学习中来, 需要公平合理地评价参与方数据在联邦合作中的贡献, 并根据贡献给予各参与方相应的回报. 给定联邦学习参与方 $N=\{1,\ldots,n\}$, 参与方贡献评估定义为计算参与方的贡献向量 $\Phi=\{\phi_1,\ldots,\phi_n\}$, 其中, ϕ_i 表示参与方 i 的数据对联邦合作的贡献大小.

评估参与方贡献需要解决如下 3 个子问题.

- 数据价值度量: 设计数据估值指标 v, 度量参与方组合 $S\subseteq N$ 的数据价值 $v(S)$, 即 $v:2^N\to\mathbb{R}$. 举例来说, 数据估值指标可以是联邦模型测试准确率: 假设参与方组合 S 对齐后的数据集为 D_S, $v(S)$代表在 D_S 上训练联邦模型并在联邦服务器测试集上评测的测试准确率. 需要注意: 数据价值度量指标评估的是参与方组合的数据价值, 而不是参与方数据对联邦合作的贡献. 第 2 节将展开调研数据估值指标.

- 参与方贡献评估: 基于数据估值指标 v, 制定贡献评估方案, 评估参与方 $i \in N$ 在联邦合作中的贡献 $\phi_i(v)$, 简写为 ϕ_i. 联邦学习参与方贡献评估是一个合作博弈问题[25], 容易想到的方案是将参与方个体的数据价值作为其对联邦的贡献, 即 $\phi_i=v(\{i\})$, 但这无法有效代表参与方对联邦合作所带来的边际价值增益, 需要探索更加公平合理的参与方贡献评估方案. 第 3 节将展开调研并比较各种参与方贡献评估方案.
- 贡献评估优化: 基于联邦学, 优化参与方贡献评估计算. 与经典的合作博弈问题不同[25], 联邦学习的数据价值度量更加复杂, 需要考虑横纵联邦设定、恶意参与方、需要训练评测模型来度量不同数据价值等问题. 第 4 节将展开调研贡献评估优化技术.

表 1 展示了 3 个参与方不同组合下, 合作训练模型的测试准确率. 其中, 参与方 1 和参与方 2 持有类似足量的基础数据, 参与方 3 有少量互补的任务关键数据. 观察到, 如果将个体价值作为贡献, 则参与方 3 的贡献会被低估; 如果考虑联邦排除和包含某个参与方的组合价值差异作为贡献, 则参与方 1 和参与方 2 因为彼此之间的可替代性, 会被认为均为零贡献. 由此可以看出, 贡献评估需要充分考虑不同参与方组合下的数据价值来综合评价. 为简化本文表述便于理解, 对全文中的常用术语作如下说明: 联邦代指联邦服务器或联邦学习平台, 参与方贡献代指参与方数据在全体参与方联邦合作训练联邦模型中的贡献, 数据估值表示数据价值的度量.

表 1 不同参与方组合数据训练联邦模型的测试准确率

S	∅	{1}	{2}	{3}	{1,2}	{1,3}	{2,3}	{1,2,3}
模型准确率(%)	50	80	80	65	80	90	90	90

2 数据估值指标

联邦学习参与方贡献评估需要解决的首要问题是如何度量数据价值, 本节综述数据估值相关指标, 为第 3 节介绍贡献评估技术奠定基础. 给定参与方组合 $S \subseteq N$, 数据估值指标 v 计算 S 对应数据的价值. 图 4 展示了数据估值指标的分类情况. 根据联邦是否提供任务的测试数据集, 数据价值度量分为测试集依赖的指标[26,27]和测试集无关的指标[28−37]. 其中, 测试集依赖的指标包含测试准确率和测试不确定性, 这类指标依赖于(完备的)测试集, 可以准确地反映数据对测试场景的价值; 测试集无关指标涵盖数据统计指标[28−32]、模型相似度[33−36]和合成价值指标[38], 这类指标适用性更广, 在特定场景或联邦无完备测试集时可用于数据估值.

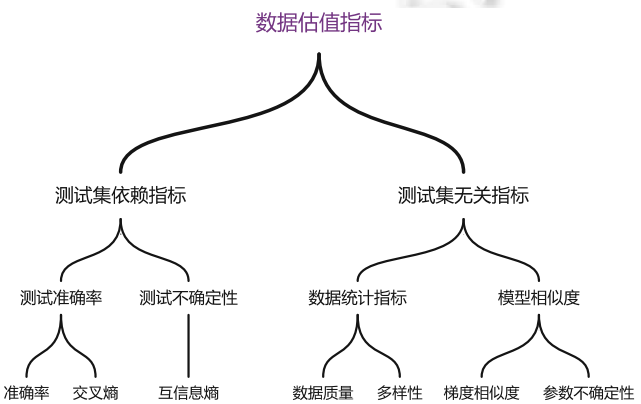

图 4 数据估值指标分类

2.1 测试集依赖指标

在理想的情况下, 联邦拥有机器学习任务的完备测试数据集, 可以基于训练数据在测试场景中的表现来度量数据价值.

2.1.1 测试准确率

最直接的方式是基于数据训练模型，通过联邦模型测试准确率来度量数据价值. 给定参与方组合 $S\subseteq N$，使用 S 中参与方数据来训练联邦模型，基于联邦测试集对训练后联邦模型进行性能评测，使用分类任务中的分类准确率或者回归任务中的测试误差来度量数据价值:

$$v(S)=\begin{cases}\dfrac{1}{|y_t|}\sum_{y_t\in y_t}\mathbb{1}[\hat{y}=y_t], & \text{分类任务}\\ -\dfrac{1}{|y_t|}\sum_{y_t\in y_t}(\hat{y}=y_t)^2, & \text{回归任务}\end{cases} \tag{1}$$

其中，y_t 是联邦测试集输出，对于每条测试结果 y_t，$\hat{y}$ 是基于 S 数据训练的联邦模型的相应预测结果. $\mathbb{1}[\cdot]$是反映分类任务预测标签与真实标签是否相等的示性函数(indicator function)，当 $\hat{y}=y_t$ 时，$\mathbb{1}[\hat{y}=y_t]=1$；否则，$\mathbb{1}[\hat{y}=y_t]=0$. 对于分类任务，当 S 训练模型预测准确率越高时，数据价值越大，组合 S 数据价值最大，$v(S)=1$. 对于回归任务，为将均方误差转换为与联邦模型性能正相关的价值度量指标，可将均方误差取相反数，当联邦模型拟合误差为 0 时价值最大，$v(S)=0$.

在实际应用中，公式(1)根据测试效果判定的标准变化而变化. 对于回归任务，公式中的均方误差可按照具体任务对单条数据样本的敏感程度修改为其他误差度量标准，比如均方根误差或者绝对误差. 对于分类任务，由于示性函数的输出只能为 0 或者 1，粗粒度地对于分类错误的测试样本同等看待，不能区分在单条测试数据上预测结果好坏的程度，因此不利于量化分类任务的具体误差. 为细粒度对正确分类置信度更低的数据设置更大的估值权重，可以基于交叉熵(cross entropy)来度量分类任务中的数据价值[26]. 在信息度量理论中，交叉熵是用来度量两个分布之间信息差异的经典模型，通过交叉熵，可以对数据价值度量做连续化处理. 以上基于测试集效果度量价值的方式不依赖于指定的联邦学习任务模型，因此可以根据实际任务需要来评估任意特定模型或者结合多种模型度量，并且基于测试准确率的价值度量同时适合于横向与纵向联邦.

2.1.2 测试不确定性

对于纵向联邦，根据不同参与方之间拥有相同数据行的特性，可以无须训练并在测试集上评测模型来度量数据价值. 最典型的是直接基于参与方的特征为测试数据标签减少的不确定性来度量数据价值，测试标签减少的不确定性越大，数据特征价值越高. 可以基于信息论的条件互信息指标来度量测试结果的不确定性[25]. 在条件互信息指标中，给定联邦测试数据$(\boldsymbol{X}_t,y_t)$，将引入参与方组合 S 的数据特征 $\boldsymbol{X}_S$ 与数据标签分布的条件互信息作为 S 的价值:

$$\begin{aligned}v(S)&=\mathbb{I}(X_S;y_t\mid X_t)\\&=\sum_{x_t\in\mathbb{X}_t}p(X_t)\sum_{x\in\mathbb{X}_S}\sum_{y_t\in\mathbb{Y}_t}p(x,y_t\mid x_t)\log\frac{p(x,y_t\mid x_t)}{p(x\mid x_t)p(y_t\mid x_t)}\\&\approx\frac{1}{|y_t|}\sum_{x_t\in\mathbb{X}_t}\sum_{x\in\mathbb{X}_S}\mathbb{N}(x,x_t,y_t)\log\frac{\mathbb{N}(x_t)\mathbb{N}(x,x_t,y_t)}{\mathbb{N}(x,x_t)\mathbb{N}(x_t,y_t)}\end{aligned} \tag{2}$$

其中，X_S, X_t, Y_t 分别代表 $\boldsymbol{X}_S$, $\boldsymbol{X}_t$, y_t 的取值空间，$\mathbb{N}(\cdot)$代表相关变量值取值的样本的频次，$\boldsymbol{X}_S$ 和 $\boldsymbol{X}_t$ 为相同数据行的不同列特征. 条件互信息通过测试集和 S 中数据的最大似然分布近似度量 S 给测试结果带来的条件互信息量. 当处理连续型特征时，需要通过分区间(partitioning)的方式对连续型属性进行离散化处理. 需要注意：虽然条件互信息亦可规避重复训练联邦模型的高昂代价，但是它依赖于对各种变量取值下样本数量的统计，这增大了泄露参与方数据信息的风险，需要更多借助差分隐私、同态加密等隐私保护方法来保证数据价值度量过程中的数据安全[27]，因此，一定程度上会牺牲价值度量的准确性或效率.

2.2 测试集无关指标

在通常的实际情况中，联邦对任务应用场景了解有限，很难获取到(完备的)测试集，此时需要设计不依赖于测试集的数据估值指标.

2.2.1 数据统计指标

数据统计指标是一类直观且易于实现的方法, 它通过数据的某些统计指标特性来度量数据的价值, 本文展开介绍自助度量与数据多样性度量指标.

- 自助度量

理想情况下, 若所有数据独立同分布, 在不具备其他知识的情况下, 认为数据价值与数据量正相关, 数据量越大, 数据价值越高[7,28], 或者根据自主报告的数据质量度量数据价值[29]. 但实际情况中, 由于各参与方持有数据来源不同, 数据收集、清洗与融合的过程也不同, 从而导致各参与方提供的数据质量也不同. 因此, 数据独立同分布的假设往往不成立. 此时, 若参与方提供的数据标签真实可靠, 则可将组合数据进行多次随机划分, 一部分作为训练集, 一部分作为测试集, 再应用第 2.1 节中的指标进行数据价值度量, 最后将多次随机划分测试结果的平均值作为组合数据价值[30]. 然而现实中, 参与方的数据标签往往未统一标准, 数据标签质量良莠不齐, 因此上述指标的度量结果可能存在偏差.

- 数据多样性度量

在不存在恶意参与方提供虚假数据时, 数据分布的多样性一定程度上可以反映数据的价值, 即数据分布多样性越高, 数据价值越大. 在联邦中, 对齐数据共有 d 个特征, 因此每一条对齐数据相当于 d 维欧式空间中的一个向量, 这些向量在空间中的分布越分散, 相当于其多样性越高. 格拉姆行列式可用于量化欧式空间内数据向量形成的平行多面体的体积[31], 格拉姆矩阵与协方差矩阵的取值呈线性关系, 当此多面体的体积越大时, 数据分布越发散, 此时数据价值越大[32]. 应用格拉姆行列式价值度量函数表示如下:

$$v(S)=\sqrt{(\boldsymbol{X}_S^T\boldsymbol{X}_S)} \tag{3}$$

其中, $\boldsymbol{X}_S\in\mathbb{R}^{|D_S|\times d}$ 表示对齐后组合 S 的数据. 在不存在恶意参与方时, 基于数据分布多样性评价数据价值对于简单的数据分布较为有效(例如高斯分布). 该度量指标的一个明显缺陷在于: 参与方可以通过大量复制自己所提供的数据, 从而增大自身数据价值度量结果. 为了应对这一攻击手段, 降低参与方大量复制自身数据所带来的风险, 一种变体的多样性度量通过简单地将 d 维欧式空间进行网格划分, 用各个子空间的数据样本均值代表该子空间, 然后再基于各子空间进行分布多样性的价值度量. 该方法同时设置衰减系数项, 当同一个子空间内数据量增多时, 衰减系数也同时增大, 从而一定程度避免数据的无限复制[32],

- 多指标复合

当联邦对于模型的应用背景足够了解, 对于所需的数据统计特征建模也足够完善时, 可以依据具体应用需求, 基于多种数据统计指标设计复合价值度量指标. 例如, 若某一特征列的取值与数据价值有直接关联, 可以直接通过计算 D_S 中该特征的取值的均值、方差等统计量来度量数据价值.

以上基于数据统计指标的价值度量指标适用于横向联邦下的数据价值度量, 它不依赖于任务、模型和测试集, 在一定程度上可以反映参与方组合的数据价值.

2.2.2 模型相似度

在假定联邦参与方的数据均对任务有利、不存在低贡献或者恶意参与方的情况下, 可以认为全体参与方参与下, 联邦合作训练的全局模型最优. 此时, 采用模型相似度指标衡量数据训练的模型与全局模型相似度, 可实现比数据统计指标面向任务更加具体准确的价值度量. 模型相似度最直接的比较方式是对比参与方组合数据训练联邦模型与联邦全局模型的参数相似度来衡量数据价值, 比如 $L2$ 范数距离, 距离越小, 则数据价值越大[33]. 但实际上, 最优模型在不同随机训练条件下, 可能有多种差异很大的模型参数能够实现类似的局部最优, 因此直接比较不合理, 优化的模型相似度价值度量方式通过不同轮次的梯度相似度或者模型参数的统计不确定性来度量数据价值.

- 梯度相似度

梯度相似度通过参与方组合 S 的数据与全体参与方合作训练全局联邦模型的梯度相似度进行比较来度量组合 S 的数据价值[34,35]. 从优化角度来看, 联邦机器学习模型的主要目标是拟合一个未知函数的近似解, 其求解方法依赖于最优化一个或多个损失函数(loss function), 以使得拟合函数尽可能地与真实的未知函数相近.

而拟合未知函数所采用的梯度在一定程度上反映了拟合优化的方向，因此，如果组合数据产生的训练梯度与假定最优梯度方向越相近，认为其与最优模型越相似，即数据价值越大. 具体而言，梯度相似度将组合 S 数据训练联邦模型的梯度更新与全体参与方数据训练全局联邦模型梯度更新之间的余弦相似度作为 S 在该轮次的贡献，聚合多轮次训练中 S 的梯度相似度，即获得 S 的数据价值:

$$v(S)=\sum_{t\in[0,iter]}\cos(\boldsymbol{u}_S^t,\boldsymbol{u}_N^t)/iter=\sum_{t\in[0,iter]}\frac{\langle \boldsymbol{u}_S^t,\boldsymbol{u}_N^t\rangle}{\|\boldsymbol{u}_S^t\|\|\boldsymbol{u}_N^t\|}\Big/iter \tag{4}$$

其中，$\boldsymbol{u}_S^t$, $\boldsymbol{u}_N^t$ 分别为组合 S 和全体参与方 N 训练联邦模型在第 t 轮机器学习训练中正则化后的梯度向量，$iter$ 表示总的梯度迭代轮次. 聚合多轮次模型梯度更新的相似度度量在某种程度上可以准确地评价参与方组合 S 与全体参与方模型的相似度. 需要注意的是，由于随机梯度下降算法、梯度剪枝和正则化的随机性，可能会存在某些轮次对数据真实价值反映不准确甚至高价值数据价值为负的情况，因此，对于高度非凸难以优化的目标函数及联邦模型对任务应用效果差等情况下，存在价值度量偏差和度量不稳定性.

- 参数不确定性

基于参与方数据对联邦任务有利的假定下，价值最高的参与方数据组合可以最大限度地优化联邦任务的目标函数，降低模型参数的不确定性，因此可以基于模型参数信息增益来度量数据价值，以避免梯度相似度指标的不稳定性，提高价值度量可靠性[36]. 基于模型参数信息增益的价值度量公式如下:

$$v(S)=\mathbb{I}(\theta;D_S)=\mathbb{H}(\theta)-\mathbb{H}(\theta;D_S),\ \mathbb{H}(\theta)=-\mathbb{E}[\log p(\theta)]=-\sum_{\theta\in\theta}p(\theta)\log p(\theta) \tag{5}$$

其中，$\mathbb{H}(\theta)$表示随机初始化下模型参数的信息熵(即不确定性)，$\mathbb{H}(\theta|D_S)$表示经过参与方组合 S 的数据 D_S 训练后模型的信息熵，两者之差度量的数据价值代表着 D_S 数据训练给模型参数减少的不确定性. 模型训练后，参数后验分布 $p(\theta|D_S)$的不确定性越小，则$\mathbb{H}(\theta|D_S)$越小，即数据价值越大. 需要注意: 参数信息增益需要度量模型参数的不确定性，即任务模型需要采用为贝叶斯统计模型. 在不确定性度量下，Lv 等人进一步实现了基于参数信息增益的数据价值度量[35]，即通过模型参数压缩的技术[37]，将深度学习模型离散化，并度量模型参数的统计不确定性. 参数信息增益相对于梯度相似度在稳定性外的另一个优势是，信息增益可以一定程度上放宽对参与方数据均对任务有利的限制，即存在少量低价值或者恶意参与方时，仍然能够根据高价值参与方组合最大限度地减少联邦任务模型参数不确定性来有效地度量数据价值. 但是，信息增益依赖于贝叶斯任务模型或者参数压缩离散化对参数的统计概率建模，当任务数据分布复杂时，或者实际应用中采用非贝叶斯机器学习模型时，信息增益的价值度量准确性可能存在偏差.

以上模型相似度相关指标在不依赖于联邦测试集的前提下，尽可能任务相关地实现了数据价值的度量. 当全部参与方数据均对联邦任务有利，或者不存在很多策略性的恶意参与方时，模型相似度能够较为准确、有效地反映数据对联邦任务的价值. 但是，当存在大量策略性的恶意参与方时，比如持有一致分布的恶意数据能够使得任务目标函数优化下模型参数不确定性最小化，仍需联邦提供测试数据来提升价值度量的有效、可靠性.

2.2.3 合成价值指标

合成价值指标是一种简化的参与方组合价值度量指标. 它通过设定某种博弈规则来给不同参与方组合赋予不同的价值. 由于合成价值指标计算简单，避免了基于真实任务和数据度量价值的高昂计算代价，它可以用于与实际任务及数据价值无关的参与方贡献评估方案或者相关优化技术性能评估. 例如，定义一个多参与方一致性表决博弈问题[38]，基于表决是否包含某个指定参与方组合来衡量参与方组合价值. 具体而言，该问题下，合成价值指标需从全体参与方 N 中随机生成一个一致表决组合 $T\subseteq N$，当参与方组合 S 中涵盖 T 时，即 $T\subseteq S$，组合 S 价值度量为 1，其他情况下价值均为 0. 合成价值函主要用于合作博弈理论中参与方贡献评估方案与相关优化技术性能评估，其优势在于高效度量参与方组合的价值，避免了理论技术的性能评估引入数据集、任务模型和训练设定等假设对技术性能验证产生的实验偏差，是一种很好的理想化技术理论验证方式.

2.3 数据估值指标对比

本节综述的数据估值指标主要围绕分类回归任务和横向或纵向联邦展开, 对于其他类型任务和横纵混合联邦相关的估值场景有待更深入的研究. 上述数据估值指标选择与联邦参与对象分类正交, 即同时适用于跨设备和跨机构联邦. 此外, 目前测试集无关的数据估值指标仍存在主观性过强, 而有效性和可靠性不够等问题, 对于不依赖于测试集或者结合不完善测试集的数据估值指标有待进一步的研究. 表 2 对比了本节综述的各种数据估值指标, 从任务、模型和测试集的依赖情况、适用的联邦类型以及优缺点方面对这些指标进行了梳理总结, 可以根据下述联邦具体场景设定来选择合适的价值度量指标.

- 当联邦能够获得有效测试数据集时, 直接通过测试准确率可以反映数据对任务测试场景最准确的价值高低.
- 当联邦无法获得(完备的)测试数据集时, 需通过数据统计指标或者训练模型相似度对数据进行侧面的价值度量. 但是当联邦中存在过多低价值或者恶意参与方时, 这些指标可能存在无法准确而有效地反映数据价值的情况.
- 纵向联邦目前只能通过测试准确率和测试不确定性来度量数据价值, 其他面向纵向联邦的不依赖于测试集的价值度量指标仍有待进一步研究.
- 当联邦希望价值度量结果具有跨模型鲁棒性时, 可以采用模型无关的价值度量指标, 基于多种任务模型来度量数据价值.

表 2 数据估值指标对比

	任务依赖	模型依赖	测试集依赖	联邦类型	优点	缺点
测试准确率[39]	√	×	√	HFL VFL	反映测试场景需要; 同时适用于横纵联邦	依赖完备测试集
测试不确定性[25]	√	×	√	VFL	无须训练评测模型; 适用于纵向联邦	依赖安全计算技术
数据统计指标[32]	×	×	×	HFL	简单直观, 不依赖于 任务、模型等上下文	不能完全反映任务 有效性, 易受攻击
梯度相似度[34]	√	×	×	HFL	反映数据一致性	易受恶意参与方影响
参数不确定性[36]	√	√	×	HFL	反映任务效果一致性	依赖于贝叶斯等离散模型

3 参与方贡献评估方案

在明确如何度量参与方组合数据价值的基础上, 本节介绍如何进一步度量参与方在联邦合作中的贡献, 具体介绍如下 4 种参与方贡献评估方案.

- 个体法[29]: 将参与方自身数据的价值度量或者相关变体作为该参与方的贡献. 个体法可以基于任何数据价值度量指标进行, 特别地, 有个体信誉[29]、个体交叉验证[26]、个体互信息[35]、个体采样[40]和个体影响函数[41]等指标. 个体法简单、高效, 未考虑参与方个体为联邦集体的价值增益, 适用于参与方数量众多的跨设备联邦场景.
- 留一法[39]: 将联邦全体中移除某个参与方造成的数据价值损失作为该参与方的贡献. 留一法仅考虑了全体组合下某个参与方带来的价值增益, 对于多个相似互可替代的参与方不公平, 适合于发掘稀缺性参与方的场景, 常被作为其他方案评测的基准方法.
- 夏普利值[39]: 枚举所有可能的参与方组合, 将参与方加入联邦的数据价值边际增益期望作为其贡献. 夏普利值方案直观、便于理解, 保证了对每个参与方个体贡献评估的公平性, 在目前联邦贡献评估中应用最为广泛.
- 最小核[42]: 将各参与方贡献估计转化为最优化问题, 其优化目标为任意参与方组合的贡献之和尽可能地大于其组合数据价值. 最小核方案设计上最优化子组合贡献分配, 保证了参与方子组合贡献评估相对于组合价值的公平性, 更加符合经济规律, 所以有利于联邦的长期稳定发展.

3.1 个体法

顾名思义, 个体法(individual)直接基于参与方自身的数据价值度量或者相关变体来评估参与方在联邦中的贡献:

$$\phi_i=v(\{i\}) \tag{6}$$

其中, ϕ_i代表参与方 i 在联邦中的贡献, $v(\{i\})$代表在价值度量指标 v 下参与方 i 的数据价值. 个体法的价值度量除了直接采用第 2 节介绍的价值度量指标之外, 还研究开发了如下基于个体信誉、交叉验证、互信息、采样实验和影响函数的专用方法.

- 个体信誉法

在假定有参与方数据价值评分的信誉系统中, 个体信誉法通过参与方历史参与联邦学习的信誉来评估参与方对联邦合作的贡献[29]. 参与方信誉可分为直接信誉和间接信誉. 直接信誉包含对参与方训练的局部模型质量测评和参与方的活跃度: 对于局部模型的质量评测可反映参与方数据的可靠性; 参与方的活跃度可以评估参与方信誉的实时有效性, 信誉值随参与活跃度降低而衰减. 间接信誉表示参与方在历史联邦学习任务中的信誉情况, 即同时考虑多个任务之间参与方信誉反馈的一致性来避免恶意欺诈. 参与方个体信誉法采用多权重的主观逻辑模型[29,43]来加权融合多个信誉评估, 综合评估参与方对联邦合作的贡献.

- 个体交叉验证法

对于需要估计参与方数据标签质量以及与联邦测试集一致性的场景中, 通过参与方本地模型和本地测试集与联邦全局模型和联邦测试集进行交叉测实验证来评估参与方对联邦合作的贡献[26]. 由于数据标注所引入的不确定性和群体偏差性, 某些参与方的数据标签可能存在着噪声. 同时, 参与方的数据分布可能与联邦测试集的数据分布存在差异, 因此, 个体交叉验证法基于交叉验证的效果来评估贡献. 具体而言, 该方法同时考虑参与方本地数据训练的模型在联邦测试集上的交叉熵, 以及联邦全局模型在参与方本地数据上的交叉熵. 交叉验证的交叉熵之和越小, 参与方数据标签噪声及与联邦测试集分布差异性越小, 参与方标签质量越高, 对联邦合作的贡献越大. 为此, 在全体参与方合作训练联邦模型的同时, 各参与方需要基于本地数据额外训练一个本地模型以计算对联邦测试集的交叉熵.

- 个体互信息法

针对希望识别低价值或者恶意参与方的场景[44], 个体互信息法通过参与方个体之间的互信息来评估参与方对联邦合作的贡献[35]. 该方法无须训练联邦模型, 每个参与方基于本地数据训练局部模型, 某个参与方对联邦的贡献是多个其他随机采样参与方的局部模型与该参与方局部模型的互信息量平均值, 互信息量越高, 则该参与方与其他参与方之间数据关联性越强, 说明该参与方提供的数据越可信. 实践中, 为了评估参与方局部模型之间的互信息, 该方法需通过模型参数压缩技术将模型参数离散化[37], 并基于模型间的相关性预测技术量化参与方模型间的互信息[45,46].

- 个体采样实验法

基于联邦学习全体参与方共同训练全局模型的设定, 可以设法直接在训练全局模型的同时, 完成各参与方的贡献评估, 以减少每个参与方额外训练模型的代价. 联邦学习训练过程中, 每轮次各参与方的数据会全部参与训练计算梯度, 如果随机改变各参与方数据的参与比例, 则相应的训练后联邦模型测试准确率会发生变化. 基于这样的观察, 可以调节不同的参与方数据参与比例, 重复训练多个联邦模型, 从而获得多个对应的测试准确率结果, 并基于深度学习技术来拟合各参与方采样率与联邦测试准确率的关系[40]. 由于联邦模型训练有多个轮次, 对于每个轮次可以重新随机采样不同比例的参与方数据量, 并获得对应轮次的测试准确率, 再通过采用序列化深度模型的方式, 拟合多轮次的测试结果来评估参与方的贡献.

基于上述拟合随机采样比例与参与方贡献之间关系的启发, 可进一步引入强化学习算法, 通过交互式的反馈, 动态地调节参与方采样比例, 使得联邦模型测试准确率快速地收敛到最高, 并将最优情况下的参与方数据被采样比例作为该参与方对联邦合作的贡献[47]. 如果某个参与方参与比例越大时联邦模型准确率越高, 则该参与方的数据贡献越大. 在此强化学习模型的定义中, 强化学习模型的状态空间是每轮次全局模型的更

新梯度, 决策空间是参与方数据被采样比例或局部梯度被采样比例, 奖励函数是强化学习决策带来的联邦测试准确率的变化. 类似地, 在没有测试集的条件下, 也可以通过强化学习的方式, 调整不同参与方的数据参与比例, 使得联邦模型的训练损失最小化, 并将最终强化学习模型收敛时各参与方最优参与比例作为参与方对联邦的贡献[48].

- 个体影响函数法

上述参与方采样比例评估法在评估参与方个体在联邦学习中的贡献时, 仍需要重复多次训练联邦模型, 以验证何种参与比例下联邦模型最优. 如果只希望训练单个联邦模型, 不再额外训练局部模型, 则可以采用影响函数技术进行参与方贡献评估[41]. 影响函数(influence function)[49]是鲁棒统计中的经典技术. 它被用来评估对训练样本的微小扰动给模型参数带来的影响, 比如量化追踪移除或者微小改变部分训练数据后对机器学习模型测试准确率带来的影响. 因此, 在联邦学习中, 可以通过评估扰动某个参与方数据后对于联邦模型性能影响的大小, 来评估该参与方对联邦的贡献. 其中, 对于线性回归模型, 可以进一步提升影响函数的计算效率, 通过二阶梯度近似技术来近似计算影响函数以高效评估参与方对联邦的贡献[50]. 需要注意的是, 影响函数在基于参与方全体数据训练联邦模型的基础上, 对模型参数进行微小扰动情况下是有效的, 比如每个参与方代表一条数据或者少量数据. 但是在参与方数量少、每个参与方持有数据量大的情况下, 扰动某个参与方会导致较大的全局联邦模型训练结果变化. 这是因为移除的数据量过大而导致的模型扰动, 此时, 影响函数无法准确地评估参与方的贡献.

个体法中, 参与方对联邦合作的贡献评估直接基于参与方自身数据价值、参与方之间对比或者参与方在联邦合作中的影响大小来评估参与方的贡献. 基于个体法简单、高效易理解地评估参与方贡献的特性, 在联邦学习参与方贡献评估相关问题中被广泛采用. 但是, 个体法未考虑到参与方组合在联邦合作中参与方对联邦带来的边际贡献情况, 因此个体法在很多情况下无法公平、有效地评估参与方对联邦的实际贡献程度. 比如, 某个参与方持有少量与其他参与方互补的数据, 能够对未被其他参与方覆盖到的关键测试场景准确率有提升, 但因为不包含基础测试场景常规数据, 导致该参与方数据单独训练局部模型的测试准确率不高, 被误判为低贡献.

3.2 留一法

留一法(leave one out)在机器学习任务中被广泛应用于交叉验证[51]. 基于留一法的思想, 可以将联邦排除掉某个参与方后的参与方组合价值边际损失作为该参与方对联邦的贡献[39]:

$$\phi_i=v(N)-v(N\backslash\{i\}) \tag{7}$$

与个体法不同, 留一法完全遵循参与方组合数据价值度量范式, 即贡献评估与数据价值度量问题正交, 可以适用于第 2 节中所有的价值度量指标. 但是, 留一法只考虑了其他参与方全部保留下某个参与方为联邦所带来的边际收益, 这种指定参与方最后加入联邦来评估贡献的方式同样存在公平性问题. 比如, 当存在多个参与方持有相同但对联邦高价值的数据时, 移除掉持有该数据的参与方中任何一个均不会对联邦测试准确率带来显著影响, 这些参与方会被评估为低价值, 但同时, 移除这些参与方将大大降低联邦的性能.

3.3 夏普利值

1953 年, 夏普利值(shapley value)被提出来以解决合作博弈问题[52]. 后来, 夏普利博弈(shapley game)被广泛用来评估联邦学习中的参与方贡献, 该方案将每个参与方的夏普利值视为该参与方的贡献[38]:

$$\phi_i=\mathbb{E}_{\pi\in\Pi}[v(S_\pi^i\cup\{i\})-v(S_\pi^i)]=\frac{1}{n!}\sum_{\pi\in\Pi}[v(S_\pi^i\cup\{i\})-v(S_\pi^i)] \tag{8}$$

其中, $\pi\in\Pi$为所有参与方的一种排列, S_π^i 为排列π中排在 i 之前的参与方组合; v 为价值函数, 如测试准确率. 参与方 i 的夏普利值可以理解为 i 在所有加入联邦次序下的边际贡献期望, 即枚举所有可能加入联邦顺序下参与方 i 给联邦带来的期望价值增益作为 i 对联邦的贡献. 夏普利值计算可以化简为如下形式:

$$\phi_i=\frac{1}{n!}\sum_{S\subseteq N\setminus\{i\}}|S|!\cdot(n-1-|S|)!\cdot[v(S\cup\{i\})-v(S)]=\frac{1}{n}\sum_{S\subseteq N\setminus\{i\}}\frac{1}{\binom{n-1}{|S|}}[v(S\cup\{i\})-v(S)] \tag{9}$$

其中，$\frac{1}{n}$ 表示 i 处于排列中任意一个位置的概率，$\binom{n-1}{|S|}$ 代表从其他参与方中选取$|S|$个参与方排列在 i 之前的组合数量. 由公式(9)可见，夏普利值计算简化为枚举所有不包含某个参与方 i 的参与组合下，引入参与方 i 所带来的边际贡献期望，计算实现上，直接枚举这些子组合数据进行估值运算，按照子组合出现概率加权求边际贡献均值即可. 由于考虑了各种潜在参与方组合计算参与方对联邦合作的边际贡献，或者说考虑了所有可能的参与方加入联邦次序，夏普利值方案与个体法和留一法相比更加公平合理，它满足如下性质[53].

- 合理性(group rationality): 联邦所有参与方均参与下的数据总价值应完全分配给所有参与方，即:

$$v(N)=\sum_{k=1}^{N}\phi_k\ .$$

- 对称性(symmetry): 两个在任何组合下边际增益均相等的参与方拥有相同的联邦贡献，比如持有相同数据，即 $v(S\cup\{i\})=v(S\cup\{j\}),\ \forall S\subseteq N\setminus\{i,j\}$，则 $\phi_i=\phi_j$.
- 零贡献(zero element): 若某个参与方在任何组合下边际贡献均为 0，不能提升提升联邦的性能，则该参与方的数据贡献为 0，即 $v(S\cup\{i\})=v(S),\ \forall S\subseteq N\setminus\{i,j\},\ \forall S\subseteq N\setminus\{i\}$，则 $\phi_i=0$.
- 可加性(additivity): 参与方的价值函数为多个指标的线性和时，参与方的价值是这多个指标评估结果的线性和，即 $\phi_i(u+v)=\phi_i(u)+\phi_i(v)$，其中，$u$ 和 v 为两个指标的价值函数.

其中，联邦合理性使得参与方贡献评估直接有效反映参与方在联邦价值指标上的贡献量，比如对联邦合作测试准确率贡献的比例. 对称性和零贡献性质常被组合使用，保证贡献评估结果客观基于价值度量指标，而不区分不同参与方[54]. 可加性保证了在线性叠加的多目标优化场景中，对于后续新添加的价值度量函数无须重新计算已完成评估的价值度量函数. 最直观的例子是测试集新增测试样本时，已测试样本无须重复参与价值度量. 夏普利值考虑了所有加入联邦次序下参与方的边际价值增益，满足了参与方间贡献评估的公平性.

3.4 最小核

上述夏普利值是合作博弈理论中众多评估方案中的一种[55]. 核方法(core)[56]是一种与夏普利值同样著名的合作博弈方案，在经济学等领域有广泛应用[57]. 根据核理论，联邦对参与方的贡献(回报)评估中，任何子组合的贡献和应大于等于该组合的数据价值. 当目标无法实现时，需要将潜在子组合中的最大亏损值最小化，即最小核方法(least core). Yan 等人分析了夏普利值法优化个体公平性的局限性，认为最小核方案保证了任何子组合获得相对于组合价值贡献应得的贡献评估，是从追求联邦长久稳定性的角度上参与方贡献评估的可行方案[42]. 具体而言，最小核方法将参与方贡献评估转化为如下的最优化问题:

$$\min\ e\ \text{s.t.}\ \begin{cases}\sum_{i\in N}\phi_i=v(N)\\ \sum_{i\in S}\phi_i+e\geqslant v(S),\ \forall S\subseteq N\end{cases} \tag{10}$$

其中，e 为最小化优化目标，即组合最大亏损值. 最小核方案最小化各组合贡献评估相对于其组合价值的亏损，优化联邦整体价值在各参与方组合分配上的平衡，即追求组合层面上的公平性. 观察公式(10)约束条件，最小核总共由 2^n 个线性不等式约束和一个等式约束组成，最小核方案的贡献求解即为经典的线性规划求解问题，比如采样一定量不等式约束后，可通过单纯形法求解最小核近似解. 对比夏普利值来看，最小核满足夏普利值除可加性外的其他性质. 此外，最小核额外满足稳定性: 任何参与方子组合的贡献和尽可能地高于其组合价值，即 $\sum_{i\in S}\phi_i+e\geqslant v(S),\ \forall S\subseteq N$. 从经济角度来看，稳定性体现了按劳分配的贡献评估(利益分配)特性，任何潜在的参与方组合均获得了至少等同于其组合数据价值的回报. 从参与方理性角度来看，任何参与方组合离开联邦均不能获得更大的回报，因此有利于联邦的长期稳定. 此外需要注意: 与夏普利值方案不同的是，最小核方案的解不唯一，即贡献分配的最优解为该线性规划问题的最优边界[56].

3.5 贡献评估方案对比

表 3 展示了本节介绍的参与方贡献评估方案对比. 个体法和留一法对参与方贡献的度量简单、高效, 贡献评估代价随着参与方数量线性增长. 但是这两种方法不满足贡献评估的公平合理性需求, 无法公平、有效地度量各参与方在联邦合作中的贡献. 比如, 它们无法兼顾考虑到不同参与方持有相似数据集或互补数据集下的评估合理性, 无法为联邦发掘出最优的参与方组合, 获取对联邦最完备有效的任务数据. 夏普利值和最小核方法均满足合理性、对称性和零贡献性质, 满足联邦合作贡献评估的公平合理性需求, 但这相应地使得贡献评估需枚举全部 2^n 种参与方组合, 评估时间代价随参与方数量呈指数增长. 从贡献评估方案与价值度量指标相关性出发, 个体法方案与价值度量指标选择依赖性强, 衍生出了基于个体信誉、影响等专用指标; 留一法、夏普利值和最小核方案与价值度量指标选择正交, 即可实现方案与价值度量指标的任意组合. 从联邦分类出发, 本节所有方案均适用于横向和纵向联邦, 由于简单、高效(线性复杂度)的优势, 个体法和留一法更适合于参与方数量众多的跨设备联邦; 同时满足公平合理性的夏普利值和最小核方案由于复杂度高, 更适合于参与方数量较少的跨机构联邦.

表 3 参与方贡献评估方案对比

	合理性	对称性	零贡献	可加性	公平性	复杂度	优点	缺点
个体法[29]	×	×	×	√	×	O(n)	简单、高效, 适用于参与方众多的跨设备联邦	个体价值评估近似, 其合作贡献未考虑个体在合作中贡献, 如稀缺性
留一法[39]	×	√	√	√	×	O(n)	适用于评估数据的稀缺性; 常作为评测基准	对数据相似的同质参与方不公平
夏普利值[39]	√	√	√	√	个体公平	O(2^n)	适用于跨机构联邦; 支持估值指标动态变化	复杂度高, 不适合跨设备联邦; 潜在联邦稳定性风险
最小核[42]	√	√	√	×	组合公平	O(2^n)	适用于跨机构联邦; 有利于联邦稳定发展	复杂度高, 不适合跨设备联邦; 不支持指标动态变化

对比夏普利值和核最小核方法来看, 主要区别是夏普利值满足可加性而不满足稳定性, 最小核不满足可加性但满足稳定性. 可加性使得夏普利值在评估后线性地新增价值度量指标时, 不再需要额外对已完成评估的价值度量指标进行重复评估, 可以大大减少重复评估代价. 当评估任务能够一次性了解全部价值度量指标时, 可以将多个价值度量指标复合成单个价值函数, 这时, 对评估方案的可加性要求是不必要的. 另一方面, 夏普利值考虑了所有加入联邦次序下参与方的边际价值增益, 满足了个体层面的公平性; 最小核保证任何参与方子组合获得相对于组合数据价值应得的贡献评估并给予相应回报激励, 满足了任意参与方组合下的公平性, 即联邦稳定性. 如果从经济角度联邦(或者数据集市)的长远发展出发, 最小核方法的稳定性特性有助于联邦的长期稳定发展需要, 避免了某些参与方子组合为追求更大收益结对离开联邦的问题.

在实际应用中, 个体法虽然不够公平合理, 但由于其直观简洁特性等原因, 仍被部分工作采纳; 而留一法主要被当作测评基准方法; 同时满足公平合理性需求的夏普利值合作博弈方案比最小核方案应用更加广泛, 是目前联邦贡献评估主要被采取的方案; 而最小核方案仍在应用的初步测评阶段.

4 贡献评估优化技术

从理论角度来看, 结合数据价值度量指标(第 2 节)与参与方贡献评估方案(第 3 节)已经可以完成联邦贡献评估. 然而在实际应用中, 公平合理的贡献评估方案往往需要穷尽枚举所有可能的参与方组合并度量不同组合的数据价值, 比如针对不同组合数据训练并评测联邦模型, 这导致评估计算产生高昂的运算代价. 此外, 联邦贡献评估还需考虑如何抵御潜在恶意参与方的影响, 避免恶意参与方影响联邦模型性能. 图 5 展示了本节介绍的针对上述需求的相关优化技术, 具体包含如下几个方面.

- 统计采样优化: 夏普利值和最小核评估方案需要枚举指数级的参与方组合来评估参与方贡献, 可以

通过统计采样，随机采样少量参与方组合来近似计算贡献，降低贡献评估计算复杂度[58−67]；在参与方数量众多的情况下，通过约束采样结果统计特性来进一步加速采样收敛[38,64]，并通过采样少量参与方进行贡献评估的方式来进一步优化效率[68].

- 联邦特性优化：联邦学习贡献评估往往需要训练并评测模型，这导致了高昂的数据价值度量代价，通过复用模型训练梯度，大大减少重复训练成本[69−72]；同时，可充分利用剪枝技术，对参与方、训练轮次和数据样本等进行剪枝，避免低效或者无效运算[73,74]. 此外，在联邦学习实践中，无法保证参与方均提供高价值无恶意数据，贡献评估技术结果可反馈作用于筛选参与方，通过按贡献调整参与度[26]、移除低贡献参与方[75]和按贡献奖励模型策略[76]来降低联邦受到低价值或者恶意参与方影响的程度.

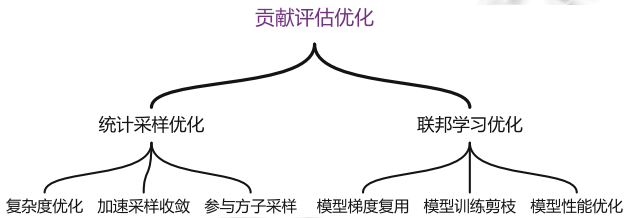

图 5 贡献评估优化技术分类

4.1 统计采样优化

4.1.1 优化计算复杂度

夏普利值和最小核评估方案在计算中需枚举 2^n 种参与方组合，在计算复杂度上属于#P-complete 问题[58]. 为了降低计算复杂度，可应用随机采样快速近似夏普利值[59]和最小核计算[60]，以实现贡献评估的(ε,δ)近似，即在大于 $1-\delta$的概率下，实现近似误差小于ε. 随机采样需要采样的参与方排列总数复杂度为 $O(n\log n)$[61]，每个排列共有 n 次数据价值计算，因此随机采样总代价为 $O(n^2\log n)$. 若假设已知参与方统计分布(在每个参与方持有单条数据时，参与方统计分布即为数据的统计分布)，夏普利值与最小核方案近似全体参与方的理论复杂度可降为 $O(n\log n)$[42,62]. 在引入分组检验这种组合搜索策略时[63]，计算理论复杂度值可进一步降至 $\sqrt{n}\log^2 n$ [54]. 具体来看，分组检验将分别包含两个不同参与方的任意参与方组合之间的价值差异的期望作为这两个参与方之间的贡献差异，通过两两对比参与方之间的贡献差异，可以构建相关约束条件下的参与方贡献最优化问题来近似计算参与方的贡献. 但是由于分组检验中分别包含不同两个参与方的两个组合中随机包含着 0 到多个不同的其他参与方，用组合间的价值差异来近似参与方之间的贡献差异的方差很大，这使得分组检验在实际应用中效果欠佳[64]. 在假设不同参与方的贡献值分布稀疏的情况下，基于压缩感知的技术可将贡献评估计算转化为凸优化问题[65]，这进一步将理论计算复杂度优化为 $O(n\log(\log n))$[54]. 对夏普利值计算的优化方法对比与分析感兴趣的读者请参考相关理论工作[66,67].

4.1.2 加速采样收敛

上述贡献评估的近似计算方法保证了理论上所需的排列样本数量复杂度上限，但实践中仍存在着采样方法收敛慢的问题，尤其是在参与方数量众多的情况下. 实践中，可以通过约束参与方排列采样结果的统计特性来进一步加速贡献近似计算的收敛速度.

- 结构化采样. 一种有效的随机采样优化策略是结构化采样. 其思想是调整随机采样样本，使得被评估的参与方出现在排列中每个位置的样本数量相等[38]. 具体而言，在近似评估某个参与方贡献时，结构化采样调整随机排列样本中该参与方出现的位置，使得其出现在排列中从前到后每个位置的排列样本数量相等，然后基于调整后的排列样本来近似该参与方的贡献.
- 引导排列采样. 结构化采样进一步演化出了引导排列采样，相比之下，引导排列采样不直接调整排列采样结果中参与方出现在各个位置的占比，而是直接约束随机采样时各参与方出现在靠前位置的频次，保证每个参与方在小样本采样中仍出现在排列靠前位置的频次相等，借此实现了贡献评估近似

计算的快速收敛[64]. 具体来说, 引导排列采样在固定参与方顺序下, 循环选定连续的 m 个参与方作为排列采样中靠前的参与方, 其中, $m<<n$; 随后, 将这 m 个靠前的参与方和其他靠后参与方分别随机排列即完成一次引导排列采样.

上述两种约束排列采样的方法均可以优化小样本下的采样结果, 保证不同参与方出现在排列中前后不同位置的分布更加均匀, 使其更快地收敛于参与方的贡献真值, 有效地降低近似计算所需的排列样本数量.

4.1.3 参与方子采样

在参与方数量庞大的场景下, 比如每条任务数据视为一个参与方或者每个终端设备的数据作为一个参与方时, 上述多项式复杂度的优化与加速排列采样收敛的方法仍需采样大量排列样本来实现较为准确的贡献评估. 为了进一步优化计算效率, 仅采样少量参与方进行联邦参与方贡献评估来优化评估效率, 并通过回归拟合的方式评估未被采样的参与方的贡献[62]. 参与方子采样方法将所有参与方近似看作参与方的统计分布, 从中随机采样少量有代表性的参与方评估其贡献, 基于这些参与方贡献评估结果来训练回归模型拟合其他参与方的贡献. 具体而言, 假设参与方的统计分布为 D, 被随机采样的参与方集合为 N, 对于任意参与方 $i\in N$, 基于夏普利方案的贡献评估公式如下:

$$\phi_i=\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}_{S\sim D^{k-1}}[v(S\cup\{i\})-v(S)] \tag{11}$$

其中, $S\sim D^{k-1}$ 代表从参与方分布中随机采样 $k-1$ 个参与方的组合 S, 即替换了经典夏普利值方案中组合 $S\subseteq N$ 的设定. 基于参与方子采样 n 个参与方来评估贡献的方式, 大大减小了实际中贡献评估的参与方数量, 并通过从分布中采样组合 S 来计算参与方边际贡献的方式, 提高了评估结果的稳定性, 即替换任何一个参与方不会对其他参与方的贡献评估产生显著影响. 需要注意的是, 因为参与方子采样评估中联邦合作的总价值是基于子采样参与方集合 N 计算, 因此不能准确地代表全部参与方的总价值, 此时, 近似计算的参与方贡献值为相对贡献值, 不满足贡献评估中的组合合理性. 此外, 参与方子采样中需要样本数量 n 足够大, 使得 n 个参与方能够有效地近似全体参与方的分布, 有效地代表其他未被采样参与方, 从而对未被采样的参与方贡献评估有较好的回归拟合效果. 基于上述已知参与方统计分布的假设下, 一些基础模型比如线性回归、高斯分布数据和核密度估计, 在夏普利值方案中存在解析解, 可避免训练与测试模型的代价, 将单个参与方贡献评估缩短至单位时间[68]. 总结来说, 参与方子采样优化适合于参与方样本量足够大, 能够在一定程度上近似参与方的真实分布的联邦任务场景, 即跨设备联邦学习, 比如个人终端和物联网场景中将数以万计的端侧设备作为参与方, 或者机器学习任务中将单个数据视作参与方来评估每条数据样本对训练模型的贡献, 但不适合只有少量参与方的联邦任务场景, 比如企业之间合作训练联邦模型的场景.

4.2 联邦学习优化

4.2.1 模型复用

贡献评估往往需要计算不同参与方组合的数据价值, 然而模型相关的价值度量指标, 比如测试准确率, 需要基于数据重新训练并评测模型, 这导致了高昂的数据价值度量代价. 为了避免重复训练联邦模型的代价, 考虑复用全体参与方组合下训练联邦模型时各参与方的梯度更新, 避免在其他参与方子组合下训练模型时各参与方重复的梯度计算, 大大减少模型训练相关的代价[69]. 模型梯度复用的直接方式是: 对于每个参与方子组合, 复用相关参与方的梯度更新来完成模型的多轮次训练, 然后评测模型度量子组合数据价值. 但是因为复用的梯度值不能代表参与方子组合数据的最优梯度方向, 经过多轮次训练后, 模型训练的累积误差大, 导致近似价值度量不够准确. 因此, 更合适的思路是评估各参与方在每轮次训练的贡献, 聚合多轮贡献来评估参与方在联邦合作中的贡献[69]. 具体而言, 全体参与方合作训练联邦模型, 在每轮次中, 根据本地数据向联邦发送梯度更新, 联邦枚举不同参与方组合梯度计算各参与方在本轮次的贡献期望, 然后, 联邦聚合所有参与方梯度完成本轮次的全局模型更新. 基于每轮次评估参与方贡献的设定, 实现了在梯度复用的同时, 大大提升评估的准确性. 但是由于每轮次均需要评测模型性能, 评测次数随训练轮次线性增大, 加大了模型评测的代价. 为了降低上述方法每轮次贡献评估的计算代价, Wang 等人在每轮次评估中仅采样部分参与方进行贡

献评估，并将未被采样的参与方在该轮次的贡献视为 0[70]. 然而，由于模型训练的性能提升增益随训练轮次逐渐收敛，该方法对于未在靠前轮次中被采样到的参与方不公平.

为了提升每轮次仅采样部分参与方进行贡献评估的公平性，观察到多轮次不同参与方组合价值构成的矩阵具有低秩特性，可以将价值度量转换为采样下的低秩矩阵补全问题[71]. 具体而言，假设有 T 轮，由于共有 2^n 个参与方组合，为了公平地评估参与方在多轮次训练中的贡献，需要求解价值度量矩阵 $V_T \in \mathbb{R}^{T\times 2^n}$ 的值，其中每个元素 $V_T[t,S]$代表组合 S 在第 t 轮的价值. 价值矩阵中，行列向列均存在相似性，其中，数据相近的不同参与方在与其他参与方组合时能够带来类似的组合价值，相邻轮次之间的价值度量存在一定相关性，从而说明价值矩阵具有低秩特性. 因此，矩阵 V_T 可基于采样补全方法来解决，即采样部分矩阵元素，将价值矩阵求解转化为优化问题 $\min\|V_T-WH\|$，其中，$W\in\mathbb{R}^{T\times r}$，$H\in\mathbb{R}^{r\times 2^n}$，其中，$r$ 为低秩矩阵 V_T 的秩，通过 W，H 近似拟合联邦训练每轮次不同组合的价值. 类似地，在纵向联邦下，基于不同参与方拥有相同行的不同列数据设定，在联邦小批量多轮次训练数据特征嵌入向量构成矩阵也具有低秩特性，可以转换成低秩矩阵补全优化问题[72].

4.2.2 模型剪枝

在联邦参与方贡献评估过程中，从采样排列中参与方的边际价值增益、模型多轮次训练的性能提升和数据样本这 3 个层面，均可以进行剪枝优化.

- 排列剪枝. 在贡献评估的随机采样优化中，对于每个排列，需要从前往后计算每个参与方加入前缀参与方组合带来的边际贡献. 在假定所有参与方的边际贡献非负的情况下，排列从前往后引入新参与方的过程中，前缀参与方组合的数据价值逐渐趋近于全体参与方组合的价值. 因此，可以设置边际增益阈值，当排列中参与方前缀组合的价值与全体参与方组合的价值差小于阈值时，往后的参与方不会带来显著边际增益，因此可以剪枝停止计算引入剩余参与方的组合数据价值，有效地提升参与方数量庞大情况下的计算效率[73].
- 训练剪枝. 机器学习模型往往需要多轮次训练才能收敛，为了降低数据估值任务中模型的训练代价，不必像模型应用性能测评中一样，尽可能地让模型收敛，在模型性能提升波动小于一定程度时进行剪枝，提早结束模型训练. 甚至为进一步提升联邦数据评估效率，可以根据任务复杂度适当提升模型学习率，仅进行单轮次模型训练[73].
- 数据剪枝. 基于局部相关特性，联邦采用 K 近邻任务模型可实现贡献评估的数据样本剪枝[74]. K 近邻方法价值计算仅关联到离测试样本最近的 k 条训练数据样本，可忽略离测试样本距离过远的数据样本，因此可以对远距离样本剪枝来提升贡献评估效率. 假设全体参与方 N 一共有$|D_N|$条数据，则 K 近邻方法下一条测试样本的 K 近邻最多有 $O|D_N|^k$ 种组合，当$|D_N|^k<<2^n$ 时，K 近邻方法可以在保证贡献评估准确性的同时，显著提升贡献评估的性能. 尤其是当每个参与方代表单条数据时，将所有数据样本根据离测试样本的距离排好序，可以直接由最远距离至最近距离样本递推计算夏普利值方案的解析解，此时，计算代价为所有数据距离排序的代价，即 $O(n\log n)$. 类似于参与方子采样优化情况，基于 K 近邻模型的优化更适合用于参与方数量大、每个参与方持有数据量少的联邦贡献评估场景.

4.2.3 模型性能优化

在联邦学习实践中，无法保证所有参与方均提供高价值无恶意的数据. 为了保证联邦学习效果，抵御恶意参与方攻击是联邦学习的一个重要研究议题[77]. 从优化联邦学习效果角度出发，贡献评估技术可用于优化联邦模型训练，即择优选择参与方数据，优化参与方参与程度，减少使用低价值或恶意参与方数据，降低恶意参与方对模型性能的负面影响[26,75,76]. 降低恶意参与方参与程度，优化联邦模型性能的主要方式如下.

- 按贡献调整联邦参与程度. 根据参与方的贡献大小，调整参与方在联邦训练中的参与程度. 在经典联邦学习中，所有参与方在联邦合作中参与程度仅与数据量挂钩，即基于 FedAvg，按照各参与方数据量加权聚合各参与方的梯度更新. 然而这种情况下，低价值或者恶意参与方持有大量数据时会影响联邦模型训练效果. 因此，可以计算每轮次中参与方的贡献大小，基于贡献加权下一轮联邦训练各参与方的参与程度. 具体而言，参与程度可以量化为参与更新联邦模型的参数比例[76]，即减小低贡献

参与方所影响到的模型参数比例. 此外, 也可以是所有参与方均参与全部模型参数更新, 联邦梯度聚合根据参与方贡献加权[26], 即降低低贡献参与方梯度对模型参数更新的影响. 需要注意的是, 当不存在完备联邦测试集时, 无法非常准确地衡量参与方对任务的贡献, 上述根据贡献细粒度调整参与方参与程度的方法不再有效.

- 设阈值移除低贡献参与方. 当不存在完备联邦测试集时, 可以根据参与方之间交叉验证来鉴定并移除低贡献参与方. 参与方交叉验证指的是每个参与方基于本地数据训练局部模型, 验证每个参与方模型在其他参与方数据上的测试准确率, 最终聚合参与方模型在其他参与方数据上验证的效果来评估该参与方的数据价值. 基于多数票决的设定, 如果某个参与方的局部模型在超过一定数量其他参与方的数据上验证效果小于某个阈值, 则认定该参与方为低价值或者恶意参与方, 将其移出联邦合作任务[75]. 需要注意的是: 低贡献参与方鉴定需要交叉验证时需要传递局部模型或者验证数据, 基于差分隐私技术对数据保护下, 可以生成安全的测试数据来传递数据实现交叉验证, 当模型规模不大时, 也可以通过传递局部模型的方式来实现.
- 按贡献奖励不同任务模型. 在很多联邦学习设定中, 参与方参与联邦学习的目的是获取性能更高的任务模型, 因此, 可以通过贡献大小来奖励相应性能的任务模型来避免低价值或者恶意参与方. 实现参与方按贡献获得不同任务模型, 需要改变经典联邦学习中每轮次训练后联邦同步最新全局模型给每个参与方的设定. 实现方法是: 联邦根据不同参与方每轮次的贡献大小, 仅反馈给该参与方相应比例数量的参与方梯度更新[76]. 在这种设定下, 在经过多轮训练后, 贡献越大的参与方收敛的本地模型与联邦服务器全局模型性能越接近, 而低贡献参与方则会相应收敛至性能更低的本地模型. 因此, 为了通过联邦合作收获性能更高的任务模型, 各参与方将会尽可能地提供更优质的训练数据. 需要注意这种方式的一个缺陷: 训练过程中, 不同参与方的模型参数更新未完全与联邦服务器全局模型同步, 因此每轮次参与方提供的梯度更新不代表全局模型更新的最优方向, 可能会导致后续轮次贡献评估结果偏差. 表 4 具体展示了本文中综述的联邦学习贡献评估方法对比.

表 4 参与方贡献评估工作对比(按年份先后排序)

	年份	联邦	价值度量	贡献评估	时间复杂度	技术方法	工作特点
Ref.[41]	2017	HFL	测试准确率	个体法	$O(n)$	影响函数	无须重复训练
SMC-Shapley[38]	2018	HFL	合成价值	夏普利值	$O(n^2\log n)$	结构化 排列采样	降低运算成本; 无需测试集
LOO, Shapley[39]	2019	HFL VFL	测试准确率	留一法 夏普利值	$O(n)$ $O(n^2\log n)$	排列采样	同时考虑横纵联邦
GTB-Shapley[54]	2019	HFL	测试准确率	夏普利值	$O(\sqrt{n}\log^2 n)$	分组检验	理论复杂度更低, 但实际近似效果差
TMC-Shapley[73]	2019	HFL	测试准确率	夏普利值	$O(n^2\log n)$	排列采样、 训练剪枝	降低运算成本
OR, MR[69]	2019	HFL	测试准确率	夏普利值	$O(n^2\log n)$	梯度复用、 多轮评估	降低运算成本; 结合多轮次训练场景
Influence[50]	2019	HFL	测试准确率	个体法	$O(n)$	影响函数、 解析解	降低计算成本; 仅适合参与方 数据量少情况
KNN-Shapley[74]	2019	HFL	测试准确率	夏普利值	$O(n^K)$	K 近邻剪枝	降低运算成本
Ref.[29]	2019	HFL	参与方信誉	个体法	$O(n)$	主观逻辑 模型	无需测试集 和模型训练
CFFL[76]	2020	HFL	测试准确率	个体法	$O(n)$	按贡献 调整参与	降低低贡献 参与方影响
FOC[26]	2020	HFL	交叉熵损失	个体法	$O(n)$	评估数据 标签质量	评估数据标签 噪声程度
FedSV[70]	2020	HFL	测试准确率	夏普利值	$O(n^2\log n)$	参与方采样、 多轮评估	降低计算成本; 结合多轮次训练场景
DVRL[48]	2020	HFL	训练损失	个体法	$O(n)$	强化学习	无需测试集

表 4 参与方贡献评估工作对比(按年份先后排序)(续)

	年份	联邦	价值度量	贡献评估	时间复杂度	技术方法	工作特点
D-Shapley[62]	2020	HFL	测试准确率	夏普利值	$O(n\log n)$	参与方 分布建模	不受参与方 动态变化影响
Ref.[36]	2020	HFL	信息增益	夏普利值	$O(n^2\log n)$	贝叶斯 统计模型	无需测试集
Least Core[42]	2021	HFL VFL	测试准确率	最小核	$O(n\log n)$	组合采样	保证组合公平, 有利于联邦稳定
RV[32]	2021	HFL	数据多样性	夏普利值	$O(n)$	格拉姆 行列式	无需测试集 和模型训练
CGSV[34]	2021	HFL	梯度相似度	个体法	$O(n)$	模型梯度 相似度	无需测试集
FedCCEA[40]	2021	HFL	测试准确率	个体法	$O(n)$	深度学习	拟合参与方 对准确率影响程度
F-RCCE[47]	2021	HFL	测试准确率	个体法	$O(n)$	强化学习	拟合参与方 准确率影响程度
GTG-Shapley[64]	2021	HFL	测试准确率	夏普利值	$O(n^2\log n)$	引导排列采样	加速收敛, 降低计算成本
FedValue[78]	2021	VFL	条件互信息	夏普利值	$O(n^2\log n)$	信息论、安全计算	保障数据安全; 无需测试集和模型训练
Fed-PCA[35]	2021	HFL	参数相似度	个体法	$O(n)$	模型参数互信息	无需测试集
Ref.[68]	2021	HFL	测试准确率	夏普利值	$O(n)$	简单模型的解析解	降低计算成本
Ref.[18]	2021	HFL	测试准确率	夏普利值	$O(\sqrt{n}\log^2 n)$	联邦去中心化	移除联邦服务器 可信假设
ComFedSV[71]	2021	HFL	测试准确率	夏普利值	$O(n^2\log n)$	参与方采样、 低秩矩阵补全	降低计算成本
VerFedSV[72]	2022	VFL	测试准确率	夏普利值	$O(n^2\log n)$	数据采样、 低秩矩阵补全	降低计算成本

如图 6 相关技术发展历程所示, 自 2016 年联邦学习框架提出以来, 参与方贡献评估相关技术得到了广泛的关注与研究. 贡献评估研究可追溯到 2017 年 Kox 等人[41]基于影响函数评估数据个体对模型准确率的影响. 在 2018 年, Campen 等人[38]首次将夏普利值方案用于评估各参与方的影响. 随后, 基于夏普利值核个体法的贡献评估技术蓬勃发展. 近两年来, Yan 等人基于最小核的贡献评估方案[42]、Fan 和 Han 等人研究了纵向联邦场景的有效估值[72,78], 这些新探索推动着贡献评估向着方案多样化、场景完善化的方向不断发展.

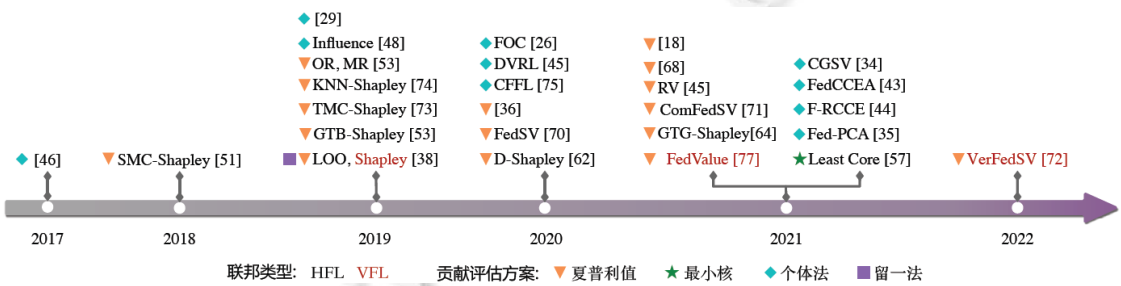

图 6 贡献评估技术发展历程

5 未来展望

参与方贡献评估作为激励参与方加入联邦的关键问题, 目前已有一些探索性研究工作, 但是现有工作在价值度量有效性和可靠性、贡献评估方案公平合理性、评估算法在联邦学习应用上的性能和安全性等问题上仍存在不足之处, 有待未来研究攻克这些挑战.

5.1 设计有效可靠的数据估值指标

目前, 相关研究已经尝试了基于测试集的测试准确率、基于信息论的信息增益、模型相似度和数据统计

特征等指标度量数据价值, 这些指标在所适应的联邦场景和价值度量的假设约束上各有优劣与不同, 未来数据估值指标设计与选择, 需从联邦场景设定和对数据与参与方的相关假设来综合考虑.

5.1.1 联邦场景设定

数据估值指标设计需明确其所适合的联邦场景, 明确价值度量在联邦场景中的泛化性和针对特定场景的性能提升之间的平衡取舍.

- 联邦类型: 明确价值度量在横向联邦、纵向联邦和横纵联邦这 3 种联邦类型的适应性. 目前, 仅有依赖测试集的测试准确率指标同时适合于横向与纵向联邦; 此外, 条件互信息指标适合于纵向联邦, 其他指标均只适用于横向联邦, 因此, 未来需要研究更多适用于纵向联邦或者适用于横纵联邦的通用指标.
- 任务相关性: 明确价值度量是否任务相关. 在明确联邦具体的分类或回归任务时, 可以针对相关任务特性的优化目标函数设计更加合理的估值指标. 而对于任务目标函数尚未完全明确或者对联邦聚合查询分析需求的任务来说, 可以考虑从数据统计指标、分布特性和时效性等角度切入, 量化联邦主观需求来设计复合指标的数据估值是未来可供探索的方向.
- 模型依赖: 明确价值度量是否依赖于具体联邦任务模型. 设计模型无关的数据估值指标可以有更大的应用场景与泛化性, 而设计任务模型依赖的指标, 比如 K 近邻法, 可以实现在特定的联邦任务下出色的性能.
- 测试集依赖: 明确价值度量是否依赖并且多大程度上依赖于联邦测试集. 目前, 价值度量完全划分为测试集依赖与测试集无关的指标, 尚未有结合不完善联邦测试集的半监督价值度量指标. 此外, 测试集无关指标仍存在价值度量不合理的问题. 未来方法需要考虑实际中如何结合联邦的不完善测试集和参与方训练集来度量价值的问题. 另外, 设计测试集无关的指标时, 需考虑价值度量与任务性能的相关合理性与潜在假设.

5.1.2 数据估值假设

设计数据估值指标需明确其对于参与方数据所服从分布的假设, 需要能够有效显著区分高低价值的数据, 在度量指标设计上需要严谨、可靠, 能够有效地抵御低价值或者恶意参与方的攻击.

- 数据分布假设: 明确数据估值指标是否依赖于参与方数据分布假设. 对联邦参与方数据分布的信息有充分了解, 能够简化数据价值度量, 比如, 假设所有参与方数据独立并来自统一分布时, 任务数据量达到能够完全反映数据统计分布规律前, 数据价值与数据量呈正相关; 达到足够完全反映数据统计分布规律后, 数据整体的价值不再随数据量升高.
- 价值度量标准: 明确数据估值指标, 能够有效地量化联邦对数据的诉求. 在联邦提供测试集场景时, 测试集的准确率即代表联邦对所需数据的诉求. 在未提供测试集或者测试集不完善的情况下, 价值度量应当反映联邦任务对数据统计指标和分布特性的需求, 不能简单凭借数据体量、分布多样性或者参与方之间的一致性来量化数据价值.
- 恶意参与方: 明确数据估值指标, 能够防御的攻击类型[77]. 实际应用中, 存在低价值、无价值或者不同类型恶意参与方的情况, 数据价值度量不能假设联邦全体参与方组合的数据价值最高, 数据估值指标设计时, 需要尽可能地甄别对任务不利的数据, 在数据中掺杂入随机噪声或者恶意数据时, 能够有效地判定数据价值下降.

5.2 探索公平合理的贡献评估方案

目前, 在联邦学习参与方贡献评估中, 夏普利值是被普遍采用的方案; 与夏普利值具备类似应用潜力的最小核法最近才被调研, 目前尚未得到充分关注与性能验证; 留一法通常被当作基准方法; 而个体法因为简单直观仍被广泛采用. 但是, 由于个体法评价基于参与方自身价值, 未从考虑参与方对联邦合作集体带来的边际贡献出发评估贡献, 因此一定程度上缺乏公平合理性. 此外, 夏普利值与最小核以外未被关注的其他合作博弈理论可能存在合适的贡献评估应用场景, 仍值得去探索. 未来贡献评估方案选择可考虑如下要点.

- 公平性: 评估方案对参与方贡献评估需要具备对称性、零贡献特性, 面向不同参与方个体或者组合的公平性, 需要充分考虑联邦场景下参与方为联邦合作带来的边际贡献, 即参与方的组合价值增益.
- 合理性: 参与方贡献评估结果中, 各参与方贡献评估之和为全体参与方联邦合作的组合数据价值, 其中可能存在某些参与方的贡献为负的情况.
- 其他性质: 评估方案具备其他有利于联邦贡献评估方案的性质, 比如价值度量指标的可加性、评估结果的稳定性等.
- 联邦场景: 明确贡献评估方案所适应的联邦场景设定. 经典的联邦学习场景是对多个地位对等的参与方的数据贡献进行评估, 然而实际中更常见的情况是参与方的出现有先后次序和不同地位权重的情况. 未来需要考虑这些新设定下的贡献评估方案探索, 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.

5.3 面向联邦学习框架的评估优化

联邦学习贡献评估问题是合作博弈理论在联邦学习主题上的延伸, 除保留了通用合作博弈问题的共性以外, 在联邦学习这个数据不动的分布式框架中有着很多问题的新特性, 比如在联邦学习场景中, 参与方组合数据价值度量代价更大, 存在数据隐私安全问题, 存在分布式环境中的同步、异步通信传输问题以及横纵联邦框架差异和后续的联邦激励等问题. 因此, 未来联邦学习贡献评估可考虑如下要点.

- 计算优化: 结合联邦学习的特性, 优化参与方贡献评估计算. 由于贡献评估往往需要验证不同参与方组合的价值, 而联邦学习多数价值度量指标需要重复训练和评测联邦模型, 因此, 如何在不损失评估准确性的前提下, 结合模型更新梯度、多轮次训练和联邦学习分布式等特性来优化贡献评估计算值得关注.
- 联邦激励: 结合其他激励要素, 综合优化联邦激励机制问题. 在完成参与方数据贡献评估后, 联邦基于参与方数据贡献、算力和参与时长等其他可量化因素, 对联邦参与方提供相应参与激励. 因此, 结合联邦参与方激励的其他因素来综合多目标优化参与方激励机制, 也是值得深入研究的方向.
- 横纵联邦: 针对横向联邦、纵向联邦与横纵混合联邦场景实现并优化贡献评估. 目前, 贡献评估研究大多分别考虑横向联邦和纵向联邦场景, 而现实中, 横纵混合联邦的情况更符合实际情况. 因此, 未来需要研究面向横纵混合联邦更通用的方案[22].
- 数据隐私安全: 联邦贡献评估需要谨慎、充分考虑方案中的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, 甚至需要移动参与方数据[21]. 因此, 未来在方案设计中, 需要研究者更多地去了解联邦数据隐私安全相关技术[79], 更加谨慎地考虑贡献评估中的数据隐私安全诉求, 避免在方案中引入难以解决的数据隐私保护弊端.

6 总 结

联邦学习框架联合不同数据持有方, 打破数据孤岛, 在保障数据安全的前提下, 赋能人工智能应用. 但是, 如何吸引高价值数据持有方加入联邦合作中来, 避免低价值、无价值和恶意参与方窃取联邦合作成果,是联邦学习首先要解决的问题. 为此, 需要制定健全的联邦学习参与方贡献评估方案, 保证参与方在联邦参与中的数据贡献得到公平合理的评估, 使参与方根据其在联邦合作中的贡献获得应得的回报, 以推动联邦学习方案落地与长效发展.

本文综述了联邦学习的参与方贡献评估技术. 本文针对参与方贡献评估面临的数据价值度量有效性和可靠性、评估方案公平合理性和如何贡献评估计算优化等问题与挑战展开了综述. 本文分别调研了数据估值指标、参与方贡献评估方案和估计计算优化这 3 个方面的技术. 本文首先综述了如何设计有效而可靠的价值度量指标, 调研了在有无联邦测试集等多种联邦场景设定下的数据价值度量指标, 其中包括测试准确率、测试不确定性和数据统计指标等. 本文接下来综述了如何设计公平合理的贡献评估方案, 介绍并分析了夏普利值

和最小核方案相对于个体法和留一法的优势, 并总结了评估方案所需的重要性质和适用场景. 此外, 本文综述了针对贡献评估固有的高计算复杂度等问题, 调研了如何进一步基于统计方法优化贡献评估计算问题、如何结合联邦学习特性来优化计算以及如何降低恶意参与方对联邦影响的问题. 最后, 本文讨论了联邦学习贡献评估目前仍面临的挑战, 并展望了未来研究工作的前进方向.

References:

[1] Economist T. The World's Most Valuable Resource is No Longer Oil, But Data. New York: The Economist, 2017.

[2] Liu Y, Fan T, Chen T, Xu Q, Yang Q. Fate: An industrial grade platform for collaborative learning with data protection. Journal of Machine Learning Research, 2021, 22(226): 1–6.

[3] Yang Q, Liu Y, Chen T, Tong Y. Federated machine learning: Concept and applications. ACM Trans. on Intelligent Systems and Technology (TIST), 2019, 10(2): 1–19.

[4] Ritzberger K, *et al*. Foundations of Non-cooperative Game Theory. Oxford University Press, 2002.

[5] Jain A, Patel H, Nagalapatti L, Gupta N, Mehta S, Guttula S, Mujumdar S, Afzal S, Sharma Mittal R, Munigala V. Overview and importance of data quality for machine learning tasks. In: Proc. of the 26th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining. 2020. 3561–3562.

[6] Wahab OA, Mourad A, Otrok H, Taleb T. Federated machine learning: Survey, multi-level classification, desirable criteria and future directions in communication and networking systems. IEEE Communications Surveys & Tutorials, 2021, 23(2): 1342–1397.

[7] Zhan Y, Zhang J, Hong Z, *et al*. A survey of incentive mechanism design for federated learning. IEEE Trans. on Emerging Topics in Computing, 2021, 10(2): 1035–1044.

[8] Zeng R, Zeng C, Wang X, Li B, Chu X. A comprehensive survey of incentive mechanism for federated learning. arXiv:2106.15406, 2021.

[9] Pei J. A survey on data pricing: From economics to data science. IEEE Trans. on Knowledge and Data Engineering, 2020, 34(10): 4586–4608.

[10] Cong Z, Luo X, Jian P, Zhu F, Zhang Y. Data pricing in machine learning pipelines. arXiv:2108.07915, 2021.

[11] Batini C, Cappiello C, Francalanci C, Maurino A. Methodologies for data quality assessment and improvement. ACM Computing Surveys (CSUR), 2009, 41(3): 1–52.

[12] Gupta N, Mujumdar S, Patel H, Masuda S, Panwar N, Bandyopadhyay S, Mehta S, Guttula S, Afzal S, Sharma Mittal R, *et al*. Data quality for machine learning tasks. In: Proc. of the 27th ACM SIGKDD Conf. on Knowledge Discovery & Data Mining. 2021. 4040–4041.

[13] Mothukuri V, Parizi RM, Pouriyeh S, Huang Y, Dehghantanha A, Srivastava G. A survey on security and privacy of federated learning. Future Generation Computer Systems, 2021, 115: 619–640.

[14] Jia R, Wu F, Sun X, Xu J, Dao D, Kailkhura B, Zhang C, Li B, Song D. Scalability vs. utility: Do we have to sacrifice one for the other in data importance quantification? In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. 2021. 8239–8247.

[15] Huang J, Talbi R, *et al*. An exploratory analysis on users' contributions in federated learning. In: Proc. of the 2nd IEEE Int'l Conf. on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA). 2020. 20–29.

[16] McMahan B, Moore E, Ramage D, Hampson S, Arcas BA. Communication-efficient learning of deep networks from decentralized data. In: Proc. of the Artificial Intelligence and Statistics. 2017. 1273–1282.

[17] Kairouz P, McMahan HB, Avent B, Bellet A, Bennis M, Bhagoji AN, Bonawitz K, Charles Z, Cormode G, Cummings R, *et al*. Advances and open problems in federated learning. Foundations and Trends® in Machine Learning, 2021, 14(1-2): 1–210.

[18] Ma S, Cao Y, Xiong L. Transparent contribution evaluation for secure federated learning on blockchain. In: Proc. of the IEEE 37th Int'l Conf. on Data Engineering Workshops (ICDEW). 2021. 88–91.

[19] Cai H, Rueckert D, Passerat-Palmbach J. 2CP: Decentralized protocols to transparently evaluate contributivity in blockchain federated learning environments. arXiv:2011.07516, 2020.

[20] Shi Y, Yu H, Leung C. A survey of fairness-aware federated learning. arXiv:2111.01872, 2021.

[21] Asad M, Moustafa A, *et al*. A critical evaluation of privacy and security threats in federated learning. Sensors, 2020, 20(24): Article No.7182. [doi: 10.3390/s20247182]

[22] Liu Y, Kang Y, Xing C, *et al*. A secure federated transfer learning framework. IEEE Intelligent Systems, 2020, 35(4): 70−82.

[23] Yang T, Andrew G, Eichner H, Sun H, Li W, Kong N, Ramage D, Beaufays F. Applied federated learning: Improving google keyboard query suggestions. arXiv:1812.02903, 2018.

[24] Zhang T, Gao L, He C, Zhang M, Krishnamachari B, Avestimehr S. Federated learning for internet of things: Applications, challenges, and opportunities. arXiv:2111.07494, 2021.

[25] Branzei R, Dimitrov D, Tijs S. Models in Cooperative Game Theory. Springer, 2008.

[26] Chen Y, Yang X, Qin X, *et al*. Dealing with label quality disparity in federated learning. In: Proc. of the Federated Learning. 2020. 108−121.

[27] Liu YX, Chen H, Liu YH, Li CP. Privacy-preserving techniques in federated learning. Ruan Jian Xue Bao/Journal of Software, 2022, 33(3): 1057−1092 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6446.htm [doi: 10.13328/j.cnki.jos.006446]

[28] Feng S, Niyato D, Wang P, Kim DI, Liang YC. Joint service pricing and cooperative relay communication for federated learning. In: Proc. of the Int'l Conf. on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). 2019. 815−820.

[29] Kang J, Xiong Z, Niyato D, Xie S, Zhang J. Incentive mechanism for reliable federated learning: A joint optimization approach to combining reputation and contract theory. IEEE Internet of Things Journal, 2019, 6(6): 10700−10714.

[30] Pandey SR, Tran NH, Bennis M, Tun YK, Manzoor A, Hong CS. A crowdsourcing framework for on-device federated learning. IEEE Trans. on Wireless Communications, 2020, 19(5): 3241−3256.

[31] Banica T, Curran S. Decomposition results for Gram matrix determinants. Journal of Mathematical Physics, 2010, 51(11): Article No.113503.

[32] Xu X, Wu Z, Foo CS, *et al*. Validation free and replication robust volume-based data valuation. In: Advances in Neural Information Processing Systems 34. 2021. 10837−10848.

[33] Zhao B, Liu X, Chen W. When crowdsensing meets federated learning: Privacy-preserving mobile crowdsensing system. arXiv: 2102.10109, 2021.

[34] Xu X, Lyu L, Ma X, *et al*. Gradient driven rewards to guarantee fairness in collaborative machine learning. In: Advances in Neural Information Processing Systems 34. 2021. 16104−16117.

[35] Lv H, Zheng Z, Luo T, Wu F, Tang S, Hua L, Jia R, Lv C. 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). 2021. 1−8.

[36] Sim RHL, Zhang Y, Chan MC, Low BKH. Collaborative machine learning with incentive-aware model rewards. In: Proc. of the Int'l Conf. on Machine Learning. 2020. 8927−8936.

[37] Agarwal N, Suresh AT, Yu FXX, Kumar S, McMahan B. CPSGD: Communication-efficient and differentially-private distributed SGD. Advances in Neural Information Processing Systems 31. 2018.

[38] Campen T, Hamers H, Husslage B, Lindelauf R. A new approximation method for the shapley value applied to the WTC 9/11 terrorist attack. Social Network Analysis and Mining, 2018, 8(1): 1−12.

[39] Wang G, Dang CX, Zhou Z. Measure contribution of participants in federated learning. In: Proc. of the IEEE Int'l Conf. on Big Data (Big Data). 2019. 2597−2604.

[40] Shyn SK, Kim D, Kim K. FedCCEA: A practical approach of client contribution evaluation for federated learning. arXiv:2106.02310, 2021.

[41] Koh PW, Liang P. Understanding black-box predictions via influence functions. In: Proc. of the Int'l Conf. on Machine Learning. PMLR, 2017. 1885−1894.

[42] Yan T, Procaccia AD. If you like shapley then you'll love the core. In: Proc. of the AAAI Conf. on Artificial Intelligence, Vol.35. 2021. 5751−5759.

[43] Jøsan g A. Subjective Logic. Cham: Springer, 2016.

[44] Lin J, Du M, Liu J. Free-riders in federated learning: Attacks and defenses. arXiv:1911.12560, 2019.

[45] Miller N, Resnick P, *et al*. Eliciting informative feedback: The peer-prediction method. Management Science, 2005, 51(9): 1359−1373.

[46] Dasgupta A, Ghosh A. Crowdsourced judgement elicitation with endogenous proficiency. In: Proc. of the 22nd Int'l Conf. on World Wide Web. 2013. 319−330.

[47] Zhao J, Zhu X, Wang J, Xiao J. Efficient client contribution evaluation for horizontal federated learning. In: Proc. of the IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP 2021). IEEE, 2021. 3060−3064.

[48] Yoon J, Arik S, Pfister T. Data valuation using reinforcement learning. In: Proc. of the Int'l Conf. on Machine Learning. PMLR, 2020. 10842−10851.

[49] Cook RD, Weisberg S. Characterizations of an empirical influence function for detecting influential cases in regression. Technometrics, 1980, 22(4):495−508.

[50] Richardson A, Filos-Ratsikas A, *et al*. Rewarding high-quality data via influence functions. arXiv:1908.11598, 2019.

[51] Kearns M, Ron D. Algorithmic stability and sanity-check bounds for leave-one-out cross-validation. Neural computation, 1999, 11(6): 1427−1453.

[52] Shapley LS. A value for *n*-person games. Annals of Mathematical Studies, 1953, 28: 307−317.

[53] Dubey P. On the uniqueness of the shapley value. Int'l Journal of Game Theory, 1975, 4(3): 131−139.

[54] Jia R, Dao D, Wang B, Hubis FA, Hynes N, Gürel NM, Li B, Zhang C, Song D, Spanos CJ. Towards efficient data valuation based on the shapley value. In: Proc. of the 22nd Int'l Conf. on Artificial Intelligence and Statistics. PMLR, 2019. 1167−1176.

[55] Peleg B, Sudhölter P. Introduction to the Theory of Cooperative Games. Springer, 2007.

[56] Schmeidler D. The nucleolus of a characteristic function game. SIAM Journal on Applied Mathematics, 1969, 17(6): 1163−1170.

[57] Telser LG. The usefulness of core theory in economics. Journal of Economic Perspectives, 1994, 8(2): 151−164.

[58] Deng X, Papadimitriou CH. On the complexity of cooperative solution concepts. Mathematics of Operations Research, 1994, 19(2): 257−266.

[59] Castro J, Gómez D, Tejada J. Polynomial calculation of the shapley value based on sampling. Computers & Operations Research, 2009, 36(5): 1726−1730.

[60] Balkanski E, Syed U, Vassilvitskii S. Statistical cost sharing. In: Advances in Neural Information Processing Systems 30. 2017.

[61] Maleki S, Tran-Thanh L, Hines G, Rahwan T, Rogers A. Bounding the estimation error of sampling-based shapley value approximation. arXiv:1306.4265, 2013.

[62] Ghorbani A, Kim M, Zou J. A distributional framework for data valuation. In: Proc. of the Int'l Conf. on Machine Learning. PMLR, 2020. 3535−3544.

[63] Du D, Hwang FK, Hwang F. Combinatorial Group Testing and Its Applications. World Scientific, 2000.

[64] Liu Z, Chen Y, Yu H, Liu Y, Cui L. GTG-shapley: Efficient and accurate participant contribution evaluation in federated learning. arXiv:2109.02053, 2021.

[65] Rauhut H. Compressive sensing and structured random matrices. In: Fornasier M, ed. Theoretical Foundations and Numerical Methods for Sparse Recovery, Vol. 9. Berlin: De Gruyter, 2010.

[66] Mitchell R, Cooper J, Frank E, *et al*. Sampling permutations for shapley value estimation. Journal of Machine Learning Research, 2022, 23(43): 1−46.

[67] Maleki S. Addressing the computational issues of the Shapley value with applications in the smart grid [Ph.D. Thesis]. University of Southampton, 2015.

[68] Kwon Y, Rivas MA, Zou J. Efficient computation and analysis of distributional shapley values. In: Proc. of the Int'l Conf. on Artificial Intelligence and Statistics. PMLR, 2021. 793−801.

[69] Song T, Tong Y, Wei S. Profit allocation for federated learning. In: Proc. of the IEEE Int'l Conf. on Big Data (Big Data). IEEE, 2019. 2577−2586.

[70] Wang T, Rausch J, Zhang C, Jia R, Song D. A principled approach to data valuation for federated learning. In: Proc. of the Federated Learning. Springer, 2020. 153−167.

[71] Fan Z, Fang H, Zhou Z, Pei J, Friedlander MP, Liu C, Zhang Y. Improving fairness for data valuation in federated learning. arXiv: 2109.09046, 2021.

[72] Fan Z, Fang H, Zhou Z, Pei J, Friedlander MP, Zhang Y. Fair and efficient contribution valuation for vertical federated learning. arXiv:2201.02658, 2022.

[73] Ghorbani A, Zou J. Data shapley: Equitable valuation of data for machine learning. In: Proc. of the Int'l Conf. on Machine Learning. PMLR, 2019. 2242−2251.

[74] Jia R, Dao D, Wang B, Hubis FA, Gurel NM, Li B, Zhang C, Spanos CJ, Song D. Efficient task-specific data valuation for nearest neighbor algorithms. arXiv:1908.08619, 2019.

[75] Lyu L, Yu J, Nandakumar K, Li Y, Ma X, Jin J, Yu H, Ng KS. Towards fair and privacy-preserving federated deep models. IEEE Trans. on Parallel and Distributed Systems, 2020, 31(11): 2524−2541.

[76] Lyu L, Xu X, Wang Q, Yu H. Collaborative fairness in federated learning. In: Proc. of the Federated Learning. Springer, 2020. 189−204.

[77] Fraboni Y, Vidal R, Lorenzi M. Free-rider attacks on model aggregation in federated learning. In: Proc. of the Int'l Conf. on Artificial Intelligence and Statistics. PMLR, 2021. 1846−1854.

[78] Han X, Wang L, Wu J. Data valuation for vertical federated learning: An information-theoretic approach. arXiv:2112.08364, 2021.

[79] Li Q, Wen Z, Wu Z, *et al*. A survey on federated learning systems: Vision, hype and reality for data privacy and protection. IEEE Trans. on Knowledge and Data Engineering, 2021. [doi: 10.1109/TKDE.2021.3124599]

附中文参考文献:

[27] 刘艺璇, 陈红, 刘宇涵, 李翠平. 联邦学习中的隐私保护技术. 软件学报, 2022, 33(3): 1057−1092. http://www.jos.org.cn/1000-9825/6446.htm [doi: 10.13328/j.cnki.jos.006446]

王勇(1996－), 男, 博士生, CCF 学生会员, 主要研究领域为联邦学习, 时空数据管理与应用.

李国良(1981－), 男, 博士, 教授, 博士生导师, CCF 杰出会员, 主要研究领域为数据库, 大数据分析和挖掘, 群体计算.

李开宇(1992－), 男, 博士, 主要研究领域为近似查询, 数据集成与众包.