面向数据交易的质量感知联邦学习贡献评价方法

田有亮[1,2,3]，支睿[1,2]，王帅[1,2]，金昆龙[1,2]

（1. 贵州大学计算机科学与技术学院，贵州 贵阳 550025；2. 贵州省密码学与区块链技术特色重点实验室，贵州 贵阳 550025；3. 贵州大学大数据与信息工程学院，贵州 贵阳 550025）

摘要：针对现有联邦学习贡献评价方法仍面临准确性不足、计算开销较高或鲁棒性受限，难以有效支撑数据交易中贡献结算与收益分配的问题，本文提出一种面向数据交易的联邦学习贡献评价方法。首先，设计具备数据质量感知能力的客户端权重优化方法，通过优化客户端聚合权重得到反映客户端质量差异的基础信号；然后，结合权重变化量、当轮边际收益与历史表现，引入主观逻辑将多源信号转化为包含正向证据、负向证据与不确定性的主观意见，实现贡献值动态映射。本文在文本、音频和图像数据集上的实验表明，所提方法在贡献评价准确性、计算效率和抗攻击鲁棒性方面均表现良好，并具备动态性与一定的可解释性。

关键词：联邦学习；贡献评价；质量感知权重优化；主观逻辑

中图分类号：TP309

文献标志码：A

doi：10.11959/j.issn.1000

A Quality-Aware Contribution Evaluation Method for Federated Learning in Data Trading

TIAN Youliang[1,2,3], ZHI Rui[1,2], WANG Shuai[1,2], JIN Kunlong[1,2]

1. College of Computer Science and Technology, Guizhou University, Guiyang 550025, Guizhou, China
2. Guizhou Provincial Key Laboratory of Cryptography and Blockchain Technology, Guizhou University, Guiyang, 550025, Guizhou, China
3. College of Big Data and Information Engineering, Guizhou University, Guiyang, 550025, Guizhou, China

Abstract: To address the problem that existing federated learning contribution evaluation methods suffer from insufficient accuracy, high computational overhead, or limited robustness, and thus cannot effectively support contribution settlement and revenue allocation in data trading, a federated learning contribution evaluation method for data trading is proposed. First, a quality-aware client weight optimization method is designed to obtain foundational signals reflecting client quality differences by optimizing client aggregation weights. Then, weight variations, per-round marginal gains, and historical performance are incorporated, and multi-source signals are transformed into subjective opinions containing positive evidence, negative evidence, and uncertainty through subjective logic, thereby enabling dynamic contribution mapping. Experiments conducted on text, audio, and image datasets demonstrate that the proposed method achieves good performance in contribution evaluation accuracy, computational efficiency, and attack robustness, while also exhibiting dynamic adaptability and interpretability.

Key words: Federated Learning, Contribution Evaluation, Quality-Aware Weight Optimization, Subjective Logic

---

收稿日期：XXXX–XX–XX；**修回日期：**XXXX–XX–XX

通信作者：支睿，gs.rzhi24@gzu.edu.cn

基金项目：国家重点研发计划项目(No. 2025YFB3109800)；国家自然科学基金项目（62272123）；贵州省高层次创新型人才项目（黔科合平台人才[2020]6008）；贵州省科技计划项目（黔科合平台人才[2020]5017，黔科合支撑[2022]一般065）。

Foundation Items： The National Natural Science Foundation of China (No.62272123), The National Key Research and Development Program of China(No. 2025YFB3109800), Project of High-level Innovative Talents of Guizhou Province under Grant([2020]6008), Science and Technology Program of Guizhou Province under Grant([2020]5017, [2022]065).

0 引言

《数字中国发展报告（2025年）》显示，我国数字经济核心产业增加值已占国内生产总值的10.5%以上，数据流通与交易需求持续增长[1]。自2014年贵阳大数据交易所成立以来，我国数据交易市场不断探索制度创新与平台化运行路径，数据要素市场化配置已形成一定实践基础[2]。然而，与传统商品交易不同，数据交易具有隐私敏感性，数据提供方难以直接开放原始数据内容[3]。随着《可信数据空间发展行动计划（2024—2028年）》等政策持续强化数据安全与隐私保护要求，传统依赖原始数据直接流转的交易模式已难以满足合规流通需求[4-5]。

联邦学习允许数据提供方在不共享原始数据的条件下共同训练全局模型，从而为隐私约束下的数据交易提供了新的技术支撑[6-7]。然而，联邦学习并未解决数据交易中的价值衡量问题。由于不同客户端在数据质量、类别分布、任务匹配度以及训练行为等方面存在显著差异，各参与方对模型性能提升的有效贡献并不相同[8]。同时，在隐私保护约束下，平台难以直接判断参与方本地数据及模型更新的质量，进而导致数据价值难定价、收益难分配等现实障碍。若缺乏有效的贡献评价方法，联邦学习仍难以支撑隐私保护下数据交易的实际落地[9]。

联邦学习贡献评价方法的优劣主要体现在准确性、效率与鲁棒性三个核心维度[10]。在数据交易场景下，还应进一步满足动态性与可解释性的要求。其中，准确性要求贡献结果能够反映客户端对全局模型的有效作用，使高质量、诚实参与的客户端在长期统计意义上获得更高贡献。效率决定方法能否实际部署。鲁棒性要求方法在噪声数据、恶意更新和复杂异质环境下仍能保持稳定可靠。动态性要求方法适应客户端随机参与、动态加入退出等训练机制。可解释性则要求贡献结果能够反映其形成依据。

现有联邦学习主流贡献评价方法可分为自报告法、基于个体性能的方法、基于夏普利值（Shapley value, SV）的方法以及基于模型相似度的方法[11]。

（1）准确性方面，自报告法依赖客户端主动披露私有信息，服务器难以验证其真实性，因此准确性易受影响。而基于个体性能的方法主要依据本地模型表现评估贡献，未能考虑联邦协作产生的贡献，因此准确性仍然受限。

（2）效率方面，基于SV的方法虽然具有较强的公平性理论基础，但需要重复评估大量客户端联盟，计算开销较高，难以满足高效性。

（3）鲁棒性方面，基于模型相似度的方法计算效率较高，但通常假设当前全局更新方向能够代表有效优化方向，在噪声数据、恶意更新或强异质环境下容易产生贡献误判，鲁棒性受到限制。

各类方法优缺点总结如表1所示。由此可见，现有方法在准确性、效率和鲁棒性等贡献评价核心维度上仍存在不同程度的不足[10]。同时，部分方法需要固定客户端参与并完成完整的联邦学习过程，难以适应联邦学习中客户端随机选择的动态训练机制。此外，多数方法缺乏对贡献形成过程的解释。总体而言，现有方法尚难满足数据交易场景对贡献评价机制的综合要求。

为此，本文提出一种面向数据交易的数据质量感知联邦学习贡献评价方法。该方法首先利用数据需求方提供的验收集计算客户端本地模型的验收集损失，并且求取该损失关于客户端聚合权重的梯度；然后通过数据质量感知权重优化算法提升高质量更新的聚合权重，抑制低质量或异常更新的聚合影响，从而获得能够反映客户端数据质量差异的优化后权重。随后，本文综合考虑权重变化、单轮边际作用与历史表现，引入主观逻辑将多源训练信号映射为客户端贡献值。

表1 各类方法优缺点总结

类别	准确性	效率	鲁棒性
自报告法	低	高	低
基于个体性能的方法	中	中	中
基于夏普利值的方法	高	低	高
基于模型相似度的方法	中	高	低

本文的具体贡献如下：

1. 提出一种新颖的质量感知联邦学习贡献评价方法。该方法通过自适应权重优化提取客户端质量基础信号，并结合主观逻辑实现贡献值映射，从而在联邦训练过程中在线评估客户端有效贡献。该方法在贡献评价准确性、效率和鲁棒性方面均取得良好效果。

2. 提出一种基于主观逻辑的客户端贡献映射方法。该方法综合考虑客户端优化权重信号、当轮边际收益及历史信息，将训练过程中的动态表现映射为客户端贡献，并在一定程度上赋予贡献评价结果可解释性。

3. 本文在文本、音频和图像数据以及对应模型上开展实验，从贡献评价准确性、效率、鲁棒性和可解释性等多个方面验证了所提方法的有效性①。

1　相关工作

在联邦学习支撑的数据交易场景中，贡献评价是实现数据定价、收益分配的基础。然而，联邦学习数据不可见的特性使平台难以检查参与方数据质量，从而增加了贡献评价的难度。围绕这一问题，现有研究提出了多类贡献评价方法，主要包括参与者自报告法、基于个体性能的方法、基于 SV 的方法以及基于模型相似度的方法。

1.1　参与者自报告法

该类方法要求客户端在上传模型更新之外，额外向服务器提供与自身数据质量、资源条件或参与成本相关的信息，服务器再依据这些信息对客户端进行质量判断。

Jiang 等[12]提出的 MCDFL 要求客户端在本地计算并回传数据质量指标，服务器再基于这些质量分数进行聚类以判断客户端的实际质量。Zeng 等[13]提出了 FMore 机制，该方法要求节点在参与前提交包含资源质量与期望支付的投标信息，服务器再依据预设评分函数对各客户端进行排序，从而进行优质参与者的筛选。

此类方法实现简单，但其有效性往往依赖客户端对额外信息的诚实反馈，且难以验证这些信息的真实性与一致性，因此在实际场景中容易受到主观申报偏差的影响，难以稳定支撑准确的贡献评价。

1.2　基于个体性能的方法

这类方法通过评估客户端本地模型或模型更新在精度、损失或收益等指标上的表现来度量其价值，并进一步据此完成奖励分配或贡献预测。

Pandey 等[14]通过引入本地相对精度作为核心量化指标，将每个客户端的个体性能与其获得的奖励直接关联。Lyu 等[15]通过服务器计算每个参与者在公共验证集上的模型更新准确率来评价其个体性能，并据此赋予声誉、分配奖励。Zhao 等[16]通过强化学习训练一个评估器，用于学习优质客户端的梯度特征，进而在后续预测客户端上传的模型更新所带来的贡献。

此类方法评价过程较为直观，但主要关注客户端个体模型的局部性能，难以刻画多客户端协同训练产生的联合收益。当某些客户端单独表现一般、但与其他客户端协作后能够提升全局模型质量时，该类方法对其协同价值的评估存在局限。

1.3　基于 SV 的方法

这类方法从合作博弈视角出发，通过评估客户端加入不同联盟前后的性能边际变化来定义贡献。

Fan 等[17]提出 FedSV，通过矩阵补全技术补全所有可能子联盟的边际贡献，从而提升联邦夏普利值的分配公平性。Liu 等[18]发现了标准 SV 在实际实施中会低估异构客户端的贡献并且为其赋予不公正的负值，并基于此提出了 CSV，通过策略等价性理论优化了效率与实际实施公平性。Lei 等[19]提出 SMC 近似算法，通过前部位置采样与后部截断来减少子联盟效用计算，从而提升估计效率。文献[18]和文献[19]均采用单轮合作博弈，将整个训练过程分解为逐轮子过程，以适应客户端随机加入与退出的动态训练场景。

但是此类方法普遍面临效率和准确性的权衡问题，或者仅当满足某强假设时能达成。例如文献[19]的近似精度对截断位置和采样规模较为敏感。而文献[18]假设客户端间存在策略等价性，并且可以通过聚类将同簇客户端视为具有同一策略；但现实场景下策略等价性假设未必成立，此时方法将退化为标准 SV。

1.4　基于模型相似度的方法

这类方法通常通过客户端本地模型更新方向与全局模型更新方向之间的关系来定义贡献，从而避免对验证集的依赖。

Wu 等[20]提出了 CoAst，利用客户端当前局部模型与后续若干轮全局更新方向之间的符号相似性衡量贡献。Jiang 等[21]提出 FedCE，利用客户端本地梯度与其余客户端聚合梯度之间的余弦差异估计贡献，认为梯度方向越不同，客户端提供的新增信

①本文公开全部实验代码与相关数据，开源地址为：https://github.com/FoxryanH/federated-contribution-evaluation.git.

息越多，贡献越大。Guo 等[22]提出 LCEFL，通过将客户端局部模型更新投影到全局模型收敛方向上，并结合夹角余弦与投影长度计算贡献值，再通过跨轮累加获得最终贡献；同时利用代表层选择降低投影计算开销。

此类方法通常具有较高效率，但其核心假设是全局更新方向能够代表有效优化方向。当存在噪声数据、异常更新或恶意攻击时，全局方向本身可能发生偏移，此时方向一致性未必等价于有效贡献，因此其准确性与鲁棒性仍存在一定局限。

尽管上述研究为联邦学习贡献评价提供了有益思路，但现有方法仍难以同时满足隐私保护数据交易场景对准确性、效率、鲁棒性、动态性与可解释性的综合要求。

进一步看，现有方法大多将贡献评价视为训练结束后的分析过程，较少从联邦聚合过程本身出发构建内生的质量感知信号，也较少将客户端单轮边际影响、历史表现与不确定性因素统一纳入贡献建模框架。基于此，本文受到客户端权重优化以及联邦学习双层优化思想的启发[23-29]，提出一种面向联邦数据交易的质量感知贡献评价方法。

相较于自报告法与基于个体性能的方法，本文基于客户端权重变化和单轮边际作用构造贡献评价依据，减少了对客户端主观申报和单独本地性能的依赖，从而提升贡献评价的准确性。相较于 SV 类方法，本文避免了大量客户端联盟枚举，从而提高了效率。相较于基于模型相似度的方法，本文利用需求方验收集构造任务导向的数据质量感知信号，降低了噪声数据和异构分布导致的贡献误判，从而增强了贡献评价的鲁棒性。此外，主观逻辑对多源信号的证据化映射也为贡献评价结果提供了一定的可解释性。据作者所知，现有研究尚未将质量感知权重优化与主观逻辑统一用于联邦学习贡献评价。

2 预备知识

2.1 联邦学习

联邦学习（Federated Learning, FL）是一种面向分布式数据场景的协同建模方法，其核心思想是在不暴露参与方原始数据的条件下，通过模型参数的交互实现联合训练，从而可以在一定程度上支撑隐私保护下的数据流通[30]。

在典型的横向 FL 场景中，系统通常由一个中心服务器和多个客户端组成。设客户端集合为 $\mathcal{U}=\{u_1,u_2,\cdots,u_N\}$，其中 N 表示客户端总数。客户端 u_i 持有本地数据集 D_i，其样本量记为 $n_i=|D_i|$，总样本量为 $n=\sum_{i=1}^{N} n_i$。FL 的目标是利用所有客户端的分布式数据，联合优化一个共享模型 $\boldsymbol{w}$，从而最小化整体经验风险：

$$\min_{\boldsymbol{w}} F(\boldsymbol{w})=\sum_{i=1}^{N}\frac{n_i}{n}F_i(\boldsymbol{w}) \tag{1}$$

其中 $F_i(\boldsymbol{w})$ 表示模型 $\boldsymbol{w}$ 在客户端 u_i 本地数据集上的损失函数。在此目标下，联邦平均（Federated Averaging, FedAvg）是最经典的聚合方法之一。设第 t 轮被选中的客户端集合为 S_t，客户端 $u_i\in S_t$ 经过本地训练后得到本地模型 $\mathbf{w}_i^t$，则服务器可按样本量进行加权聚合：

$$w^{t+1}=\sum_{i\in S_t}\alpha_i^t w_i^t \tag{2}$$

其中 $\alpha_i^t=\frac{n_i}{\sum_{j\in S_t} n_j}$ 并且有 $\alpha_i^t\geqslant 0$ 且 $\sum_{i\in S_t}\alpha_i^t=1$。

FedAvg 由于形式简单、实现高效，已成为 FL 研究中的基础聚合方法。然而在实际场景中，不同客户端的数据分布、样本规模和训练行为往往并不一致，这会导致局部更新对全局模型的作用存在明显差异[31]。同时，联邦学习通常采用部分客户端参与机制，不同轮次的参与集合动态变化，进一步增加了客户端贡献评估的复杂性。

2.2 主观逻辑

主观逻辑（Subjective Logic, SL）是一种面向不确定推理的形式化证据表示框架，适用于在证据不充分或信息存在不确定性的条件下，对命题的支持、反对与不确定状态进行建模[32]。在二元命题场景中，命题 x 的主观意见通常可表示为四元组：

$$\omega_x=(b_x,d_x,u_x,e_x) \tag{3}$$

其中 b_x 代表表示对命题 x 的信念度，d_x 表示对命题 x 的否定度；u_x 表示由于证据不足而保留下来的不确定性，e_x 表示命题 x 的基率，前三项满足归一化约束：

$$b_x+d_x+u_x=1,\quad b_x,d_x,u_x\in[0,1] \tag{4}$$

SL 中的意见通常来源于对正向证据与负向证据的统计。设支持命题 x 的正向证据量为 r_x，反对命题 x 的负向证据量为 s_x，同时引入非负先验常数

c用来控制在证据不足时不确定性所占的比例，则对应的意见可写为：

$$b_x = \frac{r_x}{r_x + s_x + c},$$
$$d_x = \frac{s_x}{r_x + s_x + c}, \tag{5}$$
$$u_x = \frac{c}{r_x + s_x + c}.$$

除了意见三元组之外，SL还引入基率e_x，用于在不确定性部分内部进行先验分配。基率通常可理解为在缺乏额外证据时，对命题x的先验偏好。基于意见三元组和基率，可以进一步定义命题x的期望值：

$$E_x = b_x + e_x u_x \tag{6}$$

该表达式说明当不确定性较低时，期望值主要由显式证据驱动；当不确定性较高时，基率将在结果中发挥更明显作用。因此，SL是一种兼具表达能力与解释能力的证据建模工具[33]。它能够为不确定环境下的判断问题提供更加细致的描述方式，这也为多源训练信号的统一表示和可解释评价提供了形式化支撑[34]。

3　系统架构

为形式化描述隐私保护约束下数据交易中的参与主体、交互关系及运行流程，本文构建一个由数据提供方、数据需求方和数据交易平台组成的联邦数据交易系统，并据此展开实体定义与系统框架说明。

3.1　实体定义

（1）数据提供方

数据提供方是参与联邦训练的客户端集合，不共享原始数据，仅在平台协调下完成模型训练并上传本地更新。由于各数据提供方在数据质量、类别分布及训练行为等方面存在差异，其对全局模型的有效贡献并不相同，部分数据提供方可能存在上传低质量更新甚至恶意更新的行为。

（2）数据需求方

数据需求方是任务发布方，自身缺乏足够的训练数据但具有明确的任务需求。需求方向平台提交任务描述，并提供一个与目标任务一致的小规模验收集D_{proxy}，该验收集不参与客户端本地训练，仅在平台侧用于贡献评价与模型质量校验。需求方有动机保证验收集的标签质量。

（3）数据交易平台

数据交易平台是联邦训练任务的服务器与交易协调者，负责接收需求方的任务与验收集、发布训练任务、组织客户端选择与全局聚合、执行贡献评价和收益结算，并在模型达标后交付给需求方。平台被视为诚实可信的中介，不拥有客户端训练数据，但可访问需求方提供的验收集。

需要指出的是，本文所引入的验收集并非传统的公共测试集假设，而是数据交易场景中由数据需求方提供的任务验收依据。相较于SV等方法需大规模测试集反复评估不同联盟模型，本文方法仅需较小规模验收集即可实现聚合优化与贡献评价，既符合交易场景的模型验收需求，也降低了评价过程对额外验证数据的依赖。

3.2　系统框架

本文面向隐私保护约束下的数据交易场景，构建由数据提供方、数据需求方和数据交易平台组成的联邦数据交易系统。系统在不暴露数据提供方原始数据的前提下，完成面向特定任务的联合模型训练，并基于训练过程中的贡献实现收益分配。系统框架如图1所示，主要包括任务提交、任务发布、本地训练与模型上传、贡献评价与优化聚合、模型交付与收益分配五个阶段。

（1）任务提交：数据需求方向数据交易平台提交任务描述与验收集。

（2）任务发布：数据交易平台初始化全局模型，并向市场中的数据提供方发布联邦训练任务，组织参与方接入训练过程。

（3）本地训练与上传模型：各数据提供方基于本地私有数据完成局部训练，并向平台上传本地模型更新。

（4）贡献评价与优化聚合：平台结合客户端上传的模型更新与数据需求方的验收集，采用本文贡献评价方法对各参与方贡献进行动态评估，并同时进行模型优化聚合。上述本地训练、贡献评价与优化聚合过程持续迭代，直至全局模型达到数据需求方目标。

（5）模型交付与收益分配：当全局模型达到预设验收要求后，平台将最终模型交付给数据需求方；随后，数据需求方支付总报酬，在泛化场景下使用模型。平台依据最终贡献评价结果完成收益分配。

综上，本文系统实现了从任务发布、隐私约束下联合训练、贡献评价到模型交付与收益结算的数据交易完整闭环。为便于后续形式化描述，系统中涉及的主要符号定义如表2所示。

表2　　符号说明

符号	定义
t	联邦训练轮次索引
s	服务器端第s次权重优化步
S_t	第t轮被选中的客户端集合
$\boldsymbol{w}^t$	第t轮全局模型参数
$\boldsymbol{w}_i^t$	客户端i在第t轮本地训练后上传的模型参数
$\boldsymbol{\alpha}^{t,s}$	第t轮第s次服务器优化时的聚合权重向量
$\alpha_i^{t,s}$	客户端i在第t轮第s次优化时的聚合权重
$\boldsymbol{\alpha}^{t,*}$	经权重优化后得到的当轮最终权重向量
$\alpha_i^{t,*}$	客户端i经权重优化后得到的当轮权重
$\bar{\boldsymbol{w}}^t(\boldsymbol{\alpha}^{t,s})$	基于当前聚合权重构造的临时聚合模型
D_{proxy}	数据需求方提供的验收集
$g_i^{t,s}$	验收集损失对客户端i聚合权重的梯度
$P_{1,i}^t, N_{1,i}^t$	客户端i的权重正证据与负证据
$P_{2,i}^t, N_{2,i}^t$	客户端i的边际作用正证据与负证据
$\hat{Q}_i^t$	客户端i在第t轮的最终贡献值

4　具体方案

4.1　具备质量感知的聚合权重优化方法

FedAvg按照客户端样本量n_i对本地模型进行加权平均，即默认样本规模越大，其更新对全局模型越重要。然而，在存在数据质量差异、局部偏移或训练噪声的联邦环境中，样本量并不能体现客户端更新对当前任务目标的实际价值。为此，本文不再仅依据样本量确定聚合权重，而是在数据需求方提供的验收集D_{proxy}上评估各客户端更新对目标任务损失的影响，并提取客户端质量的基础信号。

设第t轮被选中的客户端集合为S_t，其中$|S_t| = m$，客户端$i \in S_t$，基于当前全局模型$\boldsymbol{w}^t$完成本地训练后，上传本地模型参数$\boldsymbol{w}_i^t$，本文将客户端聚合系数显式建模为可优化向量：

$$\boldsymbol{\alpha}^{t,s} = \{\alpha_i^{t,s}\}_{i \in S_t} \tag{7}$$

其中s表示服务器侧第s次权重优化步，且满足

$$\alpha_i^{t,s} \geqslant 0, \sum_{i \in S_t} \alpha_i^{t,s} = 1 \tag{8}$$

在初始化阶段，将初始聚合权重设定为FedAvg的标准权重$\alpha_i^{t,0} = n_i / \sum_{j \in S_t} n_j$，保证方法与FedAvg在初始条件上的一致性，同时为后续质量感知优化提供稳定起点。

给定当前服务器优化步的权重向量$\boldsymbol{\alpha}^{t,s}$，服务器首先构造临时聚合模型：

$$\bar{\boldsymbol{w}}^t(\boldsymbol{\alpha}^{t,s}) = \sum_{i \in S_t} \alpha_i^{t,s} \boldsymbol{w}_i^t \tag{9}$$

随后，在数据需求方提供的代理验收集D_{proxy}上定义验收集损失函数：

$$\mathcal{L}_{proxy}^t(\boldsymbol{\alpha}^{t,s}) = \frac{1}{B_t} \sum_{b=1}^{B_t} \ell(\bar{\boldsymbol{w}}^t(\boldsymbol{\alpha}^{t,s}); \mathcal{B}_b^t) \tag{10}$$

其中$\{\mathcal{B}_b^t\}_{b=1}^{B_t}$表示代理集划分得到的$B_t$个小批次，

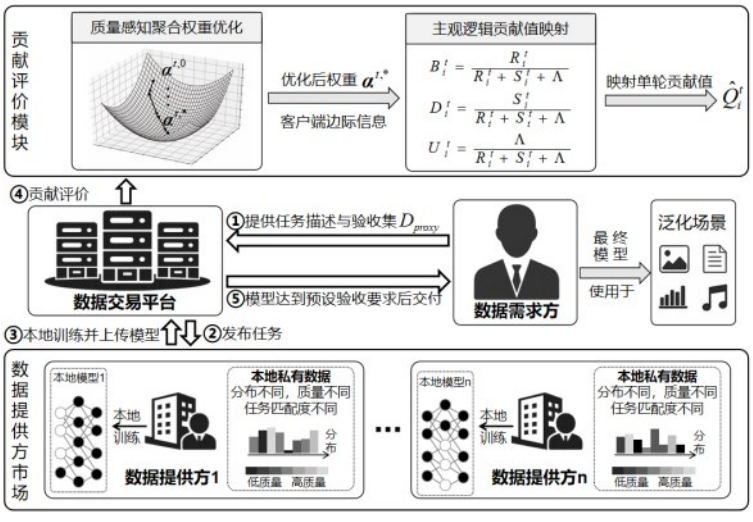

图1　系统框架

$\mathcal{B}_b^t$为第b个小批次，$\ell(\cdot)$表示交叉熵损失函数。由于本地模型$\{\boldsymbol{w}_i^t\}_{i\in S_t}$在服务器优化阶段保持固定，因此服务器实际优化的对象仅为聚合权重$\boldsymbol{\alpha}^{t,s}$。

基于上述目标，服务器计算验收集损失关于聚合权重的梯度为：

$$g_i^{t,s}=\frac{\partial \mathcal{L}_{proxy}^t(\boldsymbol{\alpha}^{t,s})}{\partial \alpha_i^{t,s}},\quad i\in S_t \tag{11}$$

该梯度反映了客户端i的聚合权重变化对当前代理任务损失的影响。当$g_i^{t,s}>0$时，增大该客户端的权重会导致验收集损失升高，说明其更新与当前任务目标的一致性相对较弱；当$g_i^{t,s}<0$时，增大其权重会降低验收集损失，表明其更新对当前任务更具正向作用。因此，$\{g_i^{t,s}\}$为质量感知的权重优化提供了的核心信号。

为避免服务器更新步长依赖固定超参数而难以适应不同轮次的客户端质量差异，本文设计了最小梯度间隙驱动的自适应服务器步长。首先记：

$$g_{\min^{t,s}}=\min_{i\in S_t} g_i^{t,s} \tag{12}$$

并定义各客户端相对于最小梯度的间隙为：

$$\Delta_i^{t,s}=g_i^{t,s}-g_{\min^{t,s}} \tag{13}$$

在此基础上，本文将未截断的服务器步长定义为：

$$\eta_{raw}^{t,s}=\sum_{i\in S_t}\Delta_i^{t,s} \tag{14}$$

并通过上界$\eta_{\max}$对其进行约束，得到实际服务器步长：

$$\eta^{t,s}=\min(\eta_{raw}^{t,s},\eta_{\max}) \tag{15}$$

其中最小梯度间隙机制仅用于构造当前轮服务器权重优化的整体步长，而不直接作为每个客户端的个体更新项。其作用在于：当各客户端代理梯度差异较大时，说明当前轮客户端质量可分性较强，服务器应采用更积极的权重重分配；当各客户端梯度接近时，说明质量差异不显著，服务器则应保持保守更新。

在获得服务器整体步长后，本文并不直接采用原始梯度进行更新，而是按梯度绝对值总量对其进行带阻尼归一化。记

$$G^{t,s}=\sum_{j\in S_t}|g_j^{t,s}| \tag{16}$$

则归一化梯度定义为：

$$\tilde{g}_i^{t,s}=\frac{g_i^{t,s}}{G^{t,s}+\lambda} \tag{17}$$

其中$\lambda>0$为阻尼系数。阻尼项的引入能够有效抑制整体梯度较小时可能出现的过强更新，从而提高服务器优化过程的数值稳定性。

在此基础上，本文采用指数乘法更新机制对聚合权重进行重分配。对于客户端$i\in S_t$，其未归一化权重更新形式为：

$$\hat{\alpha}_i^{t,s}=\alpha_i^{t,s}\exp(-\eta^{t,s}\tilde{g}_i^{t,s}) \tag{18}$$

随后再进行归一化处理，得到下一步的聚合权重：

$$\alpha_i^{t,s+1}=\frac{\hat{\alpha}_i^{t,s}}{\sum_{j\in S_t}\hat{\alpha}_j^{t,s}} \tag{19}$$

该更新方式具有两方面优势：一方面，它保证聚合权重始终非负且和为1；另一方面，它能够在连续空间中平滑地增强高质量客户端的影响并抑制低质量客户端的作用。由17式可知，$g_i^{t,s}<0$则$\tilde{g}_i^{t,s}<0$从而有$\exp(-\eta^{t,s}\tilde{g}_i^{t,s})>1$该客户端权重倾向于上升。反之，若$g_i^{t,s}>0$，则其权重倾向于下降。需要指出的是，即使当轮客户端梯度符号相同，客户端之间仍存在相对竞争关系，即梯度更小的客户端将获得更大的乘法增权效果，因此在相同初始权重条件下，其归一化后聚合权重更大。此外，随着服务器权重优化过程持续进行，客户端间梯度差异通常会减小，从而使最小梯度间隙驱动的服务器步长下降；同时，梯度绝对值整体往往减弱，使带阻尼归一化后的有效梯度幅值进一步缩小。二者共同作用，使聚合权重更新逐渐趋于平稳。

服务器按照固定步数重复上述优化。最终得到本轮优化后的聚合权重$\boldsymbol{\alpha}^{t,*}$，然后利用该权重完成当轮全局模型聚合得到下一轮的全局模型：

$$\boldsymbol{w}^{t+1}=\sum_{i\in S_t}\boldsymbol{\alpha}_i^{t,*}\boldsymbol{w}_i^t \tag{20}$$

综上，本文提出的聚合权重优化方法通过代理集损失、最小梯度间隙驱动步长、阻尼归一化和指数乘法重加权四个关键环节，实现了对客户端聚合权重的动态优化。需要强调的是，本文引入该机制的目的不仅在于改进聚合效果，更在于通过重加权过程显式区分客户端之间的质量差异。客户端相对于初始权重的变化量能够直接反映服务器对其更新价值的动态评估，因此可作为后续主观逻辑贡献映

射的核心基础信号。具备质量感知的聚合权重优化方法伪代码如算法 1 所示。

算法 1　具备质量感知的聚合权重优化方法

输入 第 t 轮选中客户端集合 S_t，各客户端本地模型 $\{w_i^t | i \in S_t\}$，初始权重 $\boldsymbol{\alpha}^{t,0}$，验收集 D_{proxy}，最大服务器步长 η_{max}，服务器优化步数 K，阻尼系数 λ

输出 第 t 轮优化后聚合权重 $\boldsymbol{\alpha}^{t,*}$

1) **for** $s = 0$ **to** $K - 1$ **do**

2) 根据当前聚合权重 $\boldsymbol{\alpha}^{t,s}$ 构造临时聚合模型 $\bar{w}^t(\boldsymbol{\alpha}^{t,s})$

3) 在验收集 D_{proxy} 上计算验收损失 $\mathcal{L}_{proxy}^t(\boldsymbol{\alpha}^{t,s})$

4) 根据式(11)计算验收损失关于聚合权重的梯度 $g_i^{t,s}$

5) 根据式(12)-(15)计算最小梯度间隙驱动的服务器步长 $\eta^{t,s}$

6) 根据式(16)-(17)对原始梯度进行带阻尼归一化，得到 $\tilde{g}_i^{t,s}$

7) 根据式(18)计算客户端权重 $\hat{\alpha}_i^{t,s} = \alpha_i^{t,s} \exp(-\eta^{t,s} \tilde{g}_i^{t,s})$

8) 根据式(19)对 $\hat{\alpha}_i^{t,s}$ 进行归一化，得到下一步的聚合权重 $\alpha_i^{t,s+1}$

9) **end for**

10) 记最终优化后聚合权重为 $\boldsymbol{\alpha}^{t,*}$

11) 利用 $\boldsymbol{\alpha}^{t,*}$ 对本轮模型进行加权聚合，得到全局模型 $\boldsymbol{w}^{(t+1)}$

12) **return** $\boldsymbol{\alpha}^{t,*}$

4.2　基于主观逻辑的客户端贡献映射方法

算法 1 通过具备质量感知的聚合权重优化方法，提取了反映客户端质量的优化后权重。然而，优化后的聚合权重本身仍主要反映服务器在当前轮次中的重加权结果，并且权重优化结果受到联邦学习当前轮次、客户端随机组合等因素的共同影响，缺乏历史稳定性与可解释性。为此，本文进一步引入主观逻辑映射模块，将训练阶段得到的权重变化信号与单轮边际作用信号统一映射为具有正负证据语义和历史先验约束的贡献值。

4.2.1　证据构造与意见建模

设客户端 i 在第 t 轮被选中，其基础聚合权重记为 $\alpha_i^{t,0}$，优化后聚合权重记为 $\alpha_i^{t,*}$。本文首先利用二者的差异提取权重侧证据。若优化后权重相对初始权重增加，则视为正证据；若其下降，则视为负证据。因此定义

$$\begin{aligned} P_{1,i}^t &= \max(\alpha_i^{t,*} - \alpha_i^{t,0}, 0) \\ N_{1,i}^t &= \max(\alpha_i^{t,0} - \alpha_i^{t,*}, 0) \end{aligned} \tag{21}$$

其中，$P_{1,i}^t$ 表示服务器在本轮优化中主动增强了客户端 i 的聚合影响，可视为权重提升正证据，$N_{1,i}^t$ 表示服务器主动削弱了其聚合影响，对应权重下降负证据。

除权重变化外，本文还引入单轮边际作用信号作为辅助证据项，具体而言，设第 t 轮在初始权重 $\boldsymbol{\alpha}^{t,0}$ 下的完整验收集损失为 $\mathcal{L}_{full}^t$ 对于客户端 i 将其对应初始权重置零并对剩余客户端权重重新归一化，可得到去除客户端 i 后的验收集损失 $\mathcal{L}_{-i}^t$。据此定义客户端 i 在第 t 轮的单轮边际作用信号为：

$$\Delta_{i,marg}^t = \mathcal{L}_{-i}^t - \mathcal{L}_{full}^t \tag{22}$$

若 $\Delta_{i,marg}^t > 0$ 说明移除客户端 i 后验收集损失升高，即该客户端对当前轮任务具有正向边际作用，若 $\Delta_{i,marg}^t < 0$ 则说明移除该客户端后验收集损失反而降低，其更新可能偏离当前任务目标。基于此，定义边际作用侧正负证据为：

$$\begin{aligned} \overline{P}_{2,i}^t &= \max(\Delta_{i,marg}^t, 0) \\ \overline{N}_{2,i}^t &= \max(-\Delta_{i,marg}^t, 0) \end{aligned} \tag{23}$$

考虑到在攻击或强异质场景下，边际作用差值的数值范围可能显著大于权重变化信号，若直接将 $\overline{P}_{2,i}^t$ 与 $\overline{N}_{2,i}^t$ 参与证据融合，容易使辅助证据主导最终评价，从而削弱服务器权重优化所提取的基础质量信号。为此，本文引入同轮量级压缩机制，对边际作用侧证据进行线性缩放。定义 $M_{P_1}^t = \max_{j \in S_t} P_{1,j}^t$，$M_{\bar{P}_2}^t = \max_{j \in S_t} \bar{P}_{2,j}^t$ 则正向边际证据调整为：

$$P_{2,i}^t = \begin{cases} \bar{P}_{2,i}^t, & M_{\bar{P}_2}^t \le M_{P_1}^t \\ \dfrac{M_{P_1}^t}{M_{\bar{P}_2}^t} \bar{P}_{2,i}^t, & M_{\bar{P}_2}^t > M_{P_1}^t \end{cases} \tag{24}$$

同理，负向边际证据调整为

$$N_{2,i}^t = \begin{cases} \bar{N}_{2,i}^t, & M_{\bar{N}_2}^t \le M_{N_1}^t \\ \dfrac{M_{N_1}^t}{M_{\bar{N}_2}^t} \bar{N}_{2,i}^t, & M_{\bar{N}_2}^t > M_{N_1}^t \end{cases} \tag{25}$$

由此，客户端 i 在第 t 轮共获得四项基础证据：权重提升正证据 P_1、单轮边际作用正证据 P_2、权

重下降负证据N_1和单轮边际作用负证据N_2。为避免单一证据源在数值上过度主导，本文对正负证据分别取均值，构造单轮正负证据总量：

$$R_i^t = \frac{P_{1,i}^t + P_{2,i}^t}{2}, S_i^t = \frac{N_{1,i}^t + N_{2,i}^t}{2} \tag{26}$$

其中，R_i^t表示客户端在当前轮中的综合正向支持强度，S_i^t表示其综合负向抑制强度。

在获得正负证据后，本文采用主观逻辑中的信念三元组对客户端状态进行建模。设主观逻辑先验质量常数为Λ，该设置使得当正负证据均较弱时，保留一定比例的不确定性。客户端i在第t轮的信任度、否定度和不确定性分别定义为

$$\begin{aligned} B_i^t &= \frac{R_i^t}{R_i^t + S_i^t + \Lambda} \\ D_i^t &= \frac{S_i^t}{R_i^t + S_i^t + \Lambda} \\ U_i^t &= \frac{\Lambda}{R_i^t + S_i^t + \Lambda} \end{aligned} \tag{27}$$

其中，B_i^t、D_i^t、U_i^t分别表示该客户端在当前轮具有正向贡献的支持程度、反对程度以及不确定性，且满足$B_i^t + D_i^t + U_i^t = 1$。

4.2.2 历史先验与贝叶斯收缩

仅依赖单轮证据容易使贡献判断受到随机波动影响，尤其在联邦训练早期，各客户端被选次数较少、服务器重加权尚不稳定，若直接根据当前轮证据输出贡献值，容易放大局部偶然性。为此，本文进一步引入基于历史优化权重的贝叶斯收缩先验。设客户端i在第t轮之前共被选中k_i^{t-1}次，其中$\tau < t$表示历史训练轮次。其历史优化权重累计和为$\sum_{\tau<t, i\in S_\tau} \alpha_i^{\tau,*}$给定伪计数超参数$M > 0$，定义客户端$i$在第$t$轮的历史平滑权重为：

$$H_i^t = \frac{M\alpha_i^{t,0} + \sum_{\tau<t, i\in S_\tau} \alpha_i^{\tau,*}}{M + k_i^{t-1}} \tag{28}$$

当$k_i^{t-1} = 0$时，上式自然退化为$H_i^t = \alpha_i^{t,0}$该设计意味着当客户端历史观测较少时，其先验更多依赖当前初始权重；随着被选次数增加，历史优化表现逐渐主导该先验，从而提高长期质量估计的稳定性。

在此基础上，本文进一步构造不确定性基率：

$$A_i^t = \frac{\alpha_i^{t,0} + H_i^t}{2} \tag{29}$$

这表示在不确定性项的分配上，本文对当前初始权重与历史平滑权重进行等权融合，从而在短期局部信息与长期历史表现之间取得平衡。

4.2.3 贡献值映射

根据主观逻辑期望形式，客户端i在第t轮的原始期望贡献定义为：

$$Q_{i,raw}^t = B_i^t + A_i^t U_i^t \tag{30}$$

其中，第一项B_i^t表示显式正证据对贡献值的直接支撑，第二项$A_i^t U_i^t$则表示在不确定条件下，由初始权重与历史先验共同分配给客户端的保守贡献。即当轮次证据充分时，贡献主要由正负证据差异决定；当证据不足时，贡献值会更多回退到由初始权重和历史表现共同反映的先验水平。因此，$Q_{i,raw}^t$并不是单轮的瞬时评分，而是在证据和历史行为共同作用下的贡献估计。

尽管原始期望贡献已经引入了不确定性与历史先验，但在训练早期，客户端间证据仍可能不稳定。为抑制早期波动对贡献评估的影响，本文进一步引入基于轮次的收缩机制。设收缩阈值为T_0，定义第t轮的收缩系数为$\rho_t = \min\left(\frac{t}{T_0}, 1\right)$，则客户端$i$在第$t$轮的最终期望贡献定义为：

$$Q_i^t = \alpha_i^{t,0} + \rho_t (Q_{i,raw}^t - \alpha_i^{t,0}) \tag{31}$$

该式表明，在联邦训练早期，当$t < T_0$时，最终贡献值将向初始权重收缩；随着轮次增加，ρ_t逐步逼近1，系统才允许贡献值更充分地反映由证据和历史先验共同决定的质量差异。该机制能够有效避免前期个别轮次的异常信号对最终贡献评估造成过大影响。

为得到第轮t内部可比较的贡献分布，本文对所有被选客户端的最终期望贡献进行归一化，定义客户端i在第t轮的归一化贡献为：

$$\hat{Q}_i^t = \frac{Q_i^t}{\sum_{j\in S_t} Q_j^t} \tag{32}$$

$\hat{Q}_i^t$即为本文最终映射得到的单轮客户端贡献值。基于主观逻辑的贡献映射方法算法如算法2所示。

算法2　基于主观逻辑的贡献映射方法

输入 单轮训练信息，先验质量常数Λ，历史伪计数M，收缩阈值T_0

输出 各轮客户端归一化贡献$\hat{Q}_i^t$

1) 初始化各客户端历史累计量与历史计数

2) **for** $t = 1$ **to** T **do**

3) **for** 每个被选中客户端 $i \in S_t$ **do**

4) 根据式(21)计算权重信号证据 $P^t_{1,i}$ 和 $N^t_{1,i}$

5) 根据式(22)-(25)计算边际作用证据 $P^t_{2,i}$ 和 $N^t_{2,i}$

6) 根据式(26)-(27)对客户端状态进行建模

7) 根据式(28)计算历史平滑权重 H^t_i

8) 根据式(29)计算不确定性基率 A^t_i

9) 根据式(30)计算原始期望贡献 $Q^t_{i,raw} = B^t_i + A^t_i U^t_i$

10) 根据式(31)计算贡献 $Q^t_i = \alpha^{t,0}_i + \rho_t(Q^t_{i,raw} - \alpha^{t,0}_i)$

11) **end for**

12) 根据式(32)计算当前轮归一化贡献

13) 更新各客户端历史统计量

14) **end for**

15) **return** $\hat{Q}^t_i$

5　实验分析

5.1　实验设置

为支持实验复现，本文实验代码开源地址为：https://github. com/FoxryanH/federated-contribution-evaluation. git。实验在 1 块 NVIDIA A30 Tensor Core GPU 上进行，所有实验均使用 Python 和 PyTorch 实现。为了验证本文的有效性，本文采用文本、音频和图像数据集以及对应模型展开实验，数据集和模型如表 3 所示。

表 3　　数据集与对应模型

数据集	模型
Yahoo Answers Topics(YA)	Text-CNN
Speech Commands(SC)	Audio-CNN
CIFAR-10(CF-10)	ResNet-18

为了验证本文贡献评价方法在异构情况下的稳定性，本文设置三种数据划分方式，分别是独立同分布（IID）、Dirichletα = 1 分布以及 Dirichletα = 0.5 的分布。

为验证所提方法的有效性，本文选择四类具有代表性的贡献评价方法作为对比基线：

1)标准 SV[35]：经典合作博弈贡献评价方法。

2)LOO[36]：基于边际性能变化的直接贡献评价方法。

3)LCEFL[22]：基于模型投影思想的高效贡献评价较新代表方法。

4)DIG-FL[42]：基于梯度内积近似的高效贡献评价方法。

由于多数 SV 类改进方法主要围绕计算效率优化或特定场景适配展开，其评价结果通常仍以标准 SV 为参考，因此本文保留标准 SV 作为参考基线。

本文的主要公共实验参数如表 4 所示，其余参数根据数据模态进行适配,例如 CIFAR-10 与 Speech Commands 使用 5 轮本地训练，Yahoo Answers 使用 1 轮本地训练；CIFAR-10 与 Speech Commands 的本地优化器为 SGD，而 Yahoo Answers 使用 Adam。

表 4　　主要公共实验参数

参数	值
客户端总数	50
每轮参与客户端数	10
联邦训练轮数	100
学习率	0.01
本地批次大小	64
测试批次大小	512
数据需求方验收集大小	1000
泛化测试集大小	10000
单轮验收集批次	1
服务器优化步数	10
梯度阻尼项	5
贡献值轮次收缩阈值 T_0	5
贝叶斯收缩的伪计数 M	3
主观逻辑先验质量常数 Λ	0.3
随机种子	42

5.2　准确性分析

为了验证本文方法的贡献评价准确性，本文选取了两种场景，一种是基于常规联邦学习的随机选择场景，另一种是排除其余干扰的受控选择场景。

5.2.1　动态随机选取下准确性分析

图 2 展示了 CF-10 数据集在 IID 条件下，经过 100 轮随机客户端采样训练后，各客户端的平均贡献值。可以观察到，在客户端数据质量一致、数据分布一致的设定下，各客户端的平均贡献整体接近均匀分布。这说明，本文方法在随机采样条件下不会凭空引入额外偏置，能够在质量一致场景中保持

较好的分配公平性。

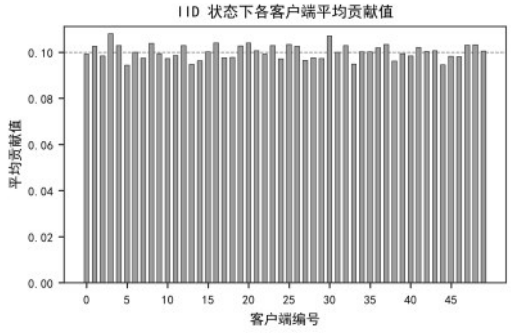

图2　IID状态下各客户端的平均贡献值

图3展示了不同数据模态与不同数据分布条件下，各类客户端的平均优化权重与平均贡献值。其中，本文设置20个干净客户端，噪声率为20%、40%、60%的客户端各10个；噪声在训练开始前通过随机标签翻转注入，并在整个训练过程中保持不变。从图中可以看出，无论是在文本、语音还是图像数据集上，也无论是在IID、Dirichletα = 1还是Dirichletα = 0.5分布下，各类客户端的平均权重与平均贡献均基本呈现出一致的单调排序关系，即干净客户端最高，其后依次为含有20%、40%和60%噪声的客户端。这一结果表明，本文方法能够较稳定地区分不同质量客户端的更新价值，并且这种区分能力不依赖于特定数据模态或特定分布设定。

图4展示了在Dirichletα = 0.5条件下不同模态，各噪声类别客户端的窗口平均权重与窗口平均贡献值。从整体趋势可以看出，在异构分布场景下，各类客户端的窗口平均贡献值整体上仍能保持与数据质量一致的排序关系。即干净客户端贡献最高，随后依次为20%、40%和60%噪声客户端。这说明，本文方法从局部训练窗口来看仍能够稳定反映不同质量客户端的相对贡献，而不会因单轮训练波动产生明显的排序偏移。值得注意的是，CF-10的窗口平均权重在训练后期出现了轻微交叉现象。这表明随着训练逐渐收敛，全局模型趋于稳定，服务器侧权重优化信号有所减弱，导致不同噪声类别之间的权重差异缩小。然而，在引入主观逻辑贡献映射后，窗口平均贡献值仍能够较好地保持与客户端质量一致的单调排序关系。该结果说明，主观逻辑映射不仅继承了服务器侧权重优化所提供的质量

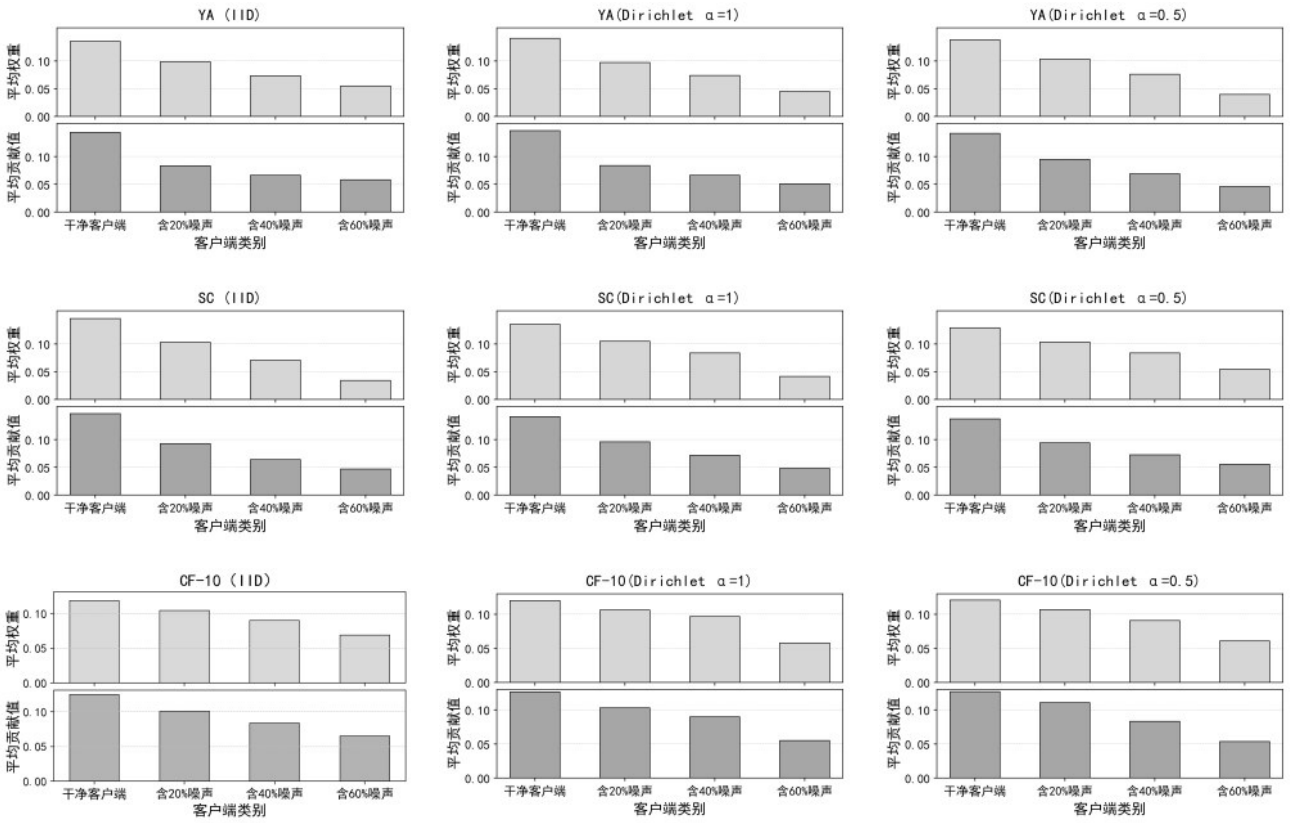

图3　不同分布不同质量情况下各类型客户端的平均权重与贡献

信号，还能够结合历史先验与单轮边际作用，对后期较弱或波动的权重信息进行平滑和修正，从而提升贡献评价结果的稳定性。

本文将客户端数据质量定义为其有效数据比例，即将质量分数记为$r_i = 1 - noise_i$。该定义描述了客户端本地数据中可为全局模型提供有效训练信号的数据样本占比，因此能够较直观地反映不同噪声客户端之间的质量差异。为评估所提方法的贡献评价准确性，本文在多种模态数据集上计算客户端最终贡献值与质量分数之间的Pearson相关系数（PCC）、Spearman秩相关系数以及单调性比例，结果如表5所示。

表5　　贡献评价准确性

数据集	分布	PCC	Spearman	单调性比例
YA	IID	0.9296	0.9551	1.0000
	a = 1	0.9267	0.9551	1.0000
	a = 0.5	0.9093	0.9400	0.9918
SC	IID	0.9691	0.9551	1.0000
	a = 1	0.9724	0.9531	0.9983
	a = 0.5	0.9480	0.9451	0.9918
CF-10	IID	0.9340	0.9275	0.9812
	a = 1	0.8916	0.9013	0.9657
	a = 0.5	0.8671	0.8405	0.9395

从结果可以看出，本文方法在三类数据集上均取得了较高的评价准确性。各设置下的Spearman相关系数和PCC均保持在较高水平，单调性比例也普遍接近1，说明本文方法得到的贡献排序与客户端有效数据质量之间具有较强的一致性。但在Dirichlet$a$ = 0.5分布下，随着数据分布异质性增强，相关性指标有所下降，尤其在CF-10数据集上更为明显，但其PCC、Spearman和单调性比例仍分别达到0.8671、0.8405和0.9395，说明本文方法在客户端随机选取且复杂异构的环境中依然具有较好的稳定性与准确性。

消融实验结果如表6所示，移除质量感知权重优化后，PCC从0.9340降至0.0090、Spearman从0.9275降至0.0634、单调性比例从0.9812降至0.6808，表明权重优化是本文方法的核心基础。移除边际贡献证据后，三项指标同步下降，说明边际证据贡献了区分信息。移除主观逻辑映射后，三项指标同样下降，表明该映射对评价值的稳定性具有重要作用。移除轮次收缩因子后，三项指标仅小幅下降，说明该机制主要抑制早期波动。消融实验结果表明，质量感知权重优化、主观逻辑映射与边际作用信号三者互为补充，共同支撑了方法在贡献评价准确性上的综合表现。

5.2.2　受控场景下的质量一致性对比

为比较不同贡献评价方法对受控数据质量差异的贡献分配表现，本文将所提方法与LCEFL、标准SV、LOO和DIG-FL进行对比[22,35-36,42]。基于客户端质量定义$r_i = 1 - noise_i$，本文构造受控场景下

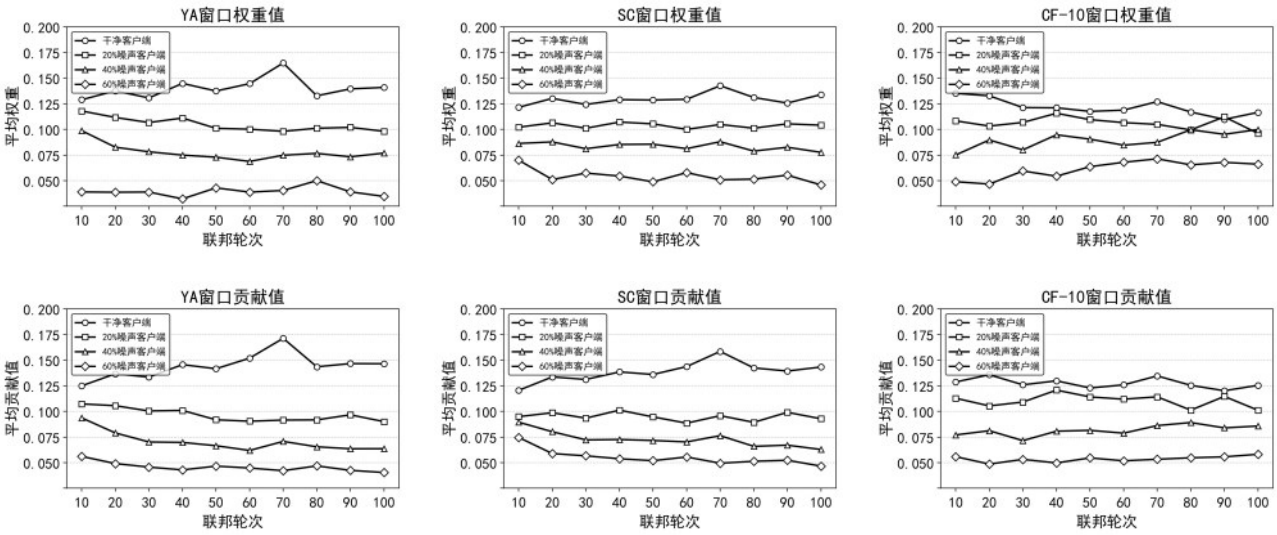

图4　异构情况下窗口平均权重与平均贡献

表 6　　消融实验

方法	PCC	Spearman	单调性比例
完整方法	0.9340	0.9275	0.9812
消融权重优化	0.0090	0.0634	0.6808
消融边际证据	0.6862	0.7264	0.8767
消融主观逻辑	0.7062	0.7350	0.8824
消融轮次收缩	0.8098	0.8441	0.9322

的质量一致性参考目标为 $Q_i^* = r_i / \left(\sum_{j=1}^{N} r_j \right)$。其中 N 表示参与客户端数。该参考目标并非一般场景下贡献评价的贡献真值，而是在训练设置保持一致的受控条件下，由客户端有效数据质量归一化得到的数据质量参考分布。本文采用 L_1 距离衡量不同方法所得贡献分布与该参考目标之间的偏差，L_1 距离越小，表示贡献评价结果与数据质量参考分布越接近。为更清晰地验证各方法的贡献分配能力，本文设计了两类受控场景：其一为同质量场景，即所有客户端数据质量相同，用于考察方法在完全均质条件下是否会产生额外偏置；其二为细粒度差异场景，即客户端噪声水平按照 0%、10%、20%、30%、40% 和 50% 逐级变化，用于检验方法对细微质量差异的分辨能力。

实验结果如表 7 所示，在同质量场景下，距离为 0.0111，说明本文方法能够为同质量客户端分配近似一致的贡献值；在细粒度差异情境下距离为 0.0417，低于 LCEFL[22]、标准 SV[35]、DIG-FL[42] 和 LOO[36]，表明所提方法所得贡献分布与受控场景下的数据质量参考目标具有更高一致性。

需要指出的是，LCEFL 在含噪场景下出现逆向排序，可能与其基于客户端更新在全局模型更新方向上的投影有关。当噪声更新已经使全局模型方向偏离实际最优方向时，干净客户端无法在已偏方向上取得更大投影；相反，噪声客户端由于局部更新幅值更大，往往会获得更高的投影值。归一化后，这种幅值优势进一步放大，最终导致质量越低的客户端反而获得更高的贡献得分。

5.3　效率分析

由表 8 可知，不同贡献评价方法在单轮时间复杂度上存在显著差异。本文方法的单轮复杂度为 $O(C_{train} + nd + Knd + n^2 d + C_{sub})$，其中 C_{train} 表示并行本地训练开销，n 为当轮参与人数，nd 表示模型聚合开销，Knd 表示服务器权重优化开销，$n^2 d$ 表示单轮边际作用计算开销，C_{sub} 表示主观逻辑贡献映射开销。由于 C_{sub} 仅涉及线性运算，且联邦学习中本地训练通常占据主要时间开销，因此本文方法的主导项仍为 $O(C_{train})$。LCEFL[22] 的单轮复杂度为 $O(C_{train} + nd_s + nd)$，d_s 表示用于代表层投影计算的参数维度，主导项同样为 $O(C_{train})$；DIG-FL[42] 的单轮复杂度为 $O(n(C_{train} + d) + C_{val})$，其中 C_{val} 表示在验证集上计算一次全模型梯度的开销。而 LOO[36] 和 SV[35] 的复杂度分别为

表 7　　受控场景下贡献评价一致性对比

场景	方法	客户端贡献值						L_1 距离
		u_1	u_2	u_3	u_4	u_5	u_6	
同质量场景	参考目标	0.1666	0.1666	0.1666	0.1666	0.1666	0.1666	0.0000
	本文方案	0.1691	0.1644	0.1640	0.1659	0.1688	0.1675	0.0111
	LCEFL[22]	0.1593	0.1721	0.1740	0.1626	0.1680	0.1636	0.0286
	SV[35]	0.1655	0.1684	0.1725	0.1655	0.1635	0.1644	0.0152
	LOO[36]	0.1616	0.1560	0.1682	0.1770	0.1722	0.1647	0.0351
	DIG-FL[42]	0.1705	0.1673	0.1885	0.1439	0.1546	0.1750	0.0694
细粒度差异场景	参考目标	0.2222	0.2000	0.1778	0.1556	0.1333	0.1111	0.0000
	本文方案	0.2251	0.1945	0.1639	0.1584	0.1317	0.1261	0.0417
	LCEFL[22]	0.0827	0.1109	0.1374	0.1989	0.2342	0.2356	0.5377
	SV[35]	0.3703	0.2828	0.1936	0.1073	0.0458	0.0000	0.4936
	LOO[36]	0.4719	0.2652	0.1879	0.0748	0.0000	0.0000	0.6502
	DIG-FL[42]	0.3782	0.2573	0.1555	0.0902	0.0703	0.0484	0.4267

$O(n(C_{train}+nd))$和$O(2^n(C_{train}+nd))$，其主导项分别为$O(nC_{train})$和$O(2^nC_{train})$。

表8　各贡献评价方案单轮时间复杂度

方案	单轮理论时间复杂度	复杂度主导项
本文方案	$O(C_{train}+nd+Knd+n^2d+C_{sub})$	$O(C_{train})$
LCEFL[22]	$O(C_{train}+nd_s+nd)$	$O(C_{train})$
SV[35]	$O(2^n(C_{train}+nd))$	$O(2^nC_{train})$
LOO[36]	$O(n(C_{train}+nd))$	$O(nC_{train})$
DIG-FL[42]	$O(n(C_{train}+d)+C_{val})$	$O(nC_{train})$

总体来看，本文方法虽然额外引入了服务器侧权重优化和单轮边际作用计算，但其主导项仍与LCEFL一样保持在$O(C_{train})$量级，明显优于LOO的线性增长和SV的指数增长。

由图5可知，本文方法与LCEFL的整体耗时始终保持在较低水平，且随客户端数量变化仅出现小幅波动，说明二者都具有较好的效率表现。相比之下，DIG-FL[42]的耗时随参与方数量呈近似线性增长，其每轮需在验证集上计算全模型梯度，引入了额外串行负载，导致实际运行时间与LOO相近，而SV则以指数趋势增长。进一步来看，本文方法的实际运行时间略低于LCEFL。其原因在于，LCEFL需要先通过联邦训练确定代表层和全局模型更新方向，再回溯计算各客户端的模型投影贡献，因此实际运行中会产生额外的记录、回退和投影计算开销。相比之下，本文方法仅需一次联邦训练，并在训练过程中完成服务器侧权重优化与边际作用分析，因此整体效率更优。本文方法在保证贡献评价能力的同时，体现了其高效性。

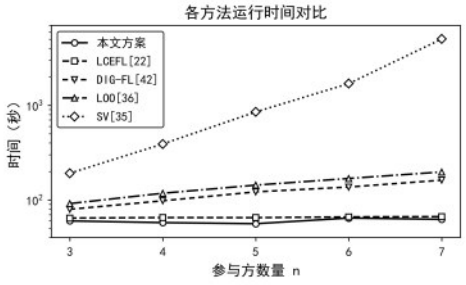

图5　单轮平均所需时间与客户端数量的关系

5.4　贡献评价鲁棒性分析

为进一步验证本文方法在复杂环境下的贡献评价鲁棒性，本文在图像数据集上构造了同时包含客户端随机选取、恶意攻击、噪声干扰和非独立同分布数据划分的实验场景，以考察所提方法在多重扰动条件下的稳定性。设置Dirichlet$a$ = 0.5，在50个客户端中，设置25个干净良性客户端、10个拜占庭恶意客户端，以及噪声率分别为20%、40%和60%的客户端各5个。对于良性噪声客户端，其质量分数定义为r_i = 1 - $noise_i$；对于恶意客户端，由于其上传更新旨在破坏全局模型，本文将其质量分数设为0。

需明确，本文并非专门设计拜占庭防御机制，因此攻击实验的目标不是抵御目前最强的自适应攻击，而是验证在典型模型投毒扰动下，所提方法是否仍能保持贡献评价能力。为此，本文选取方向反转、幅值放大和随机扰动三类典型攻击，分别对应Sign-Flipping(SF)、Scaling Attack(SA)和Gaussian Noise Injection(GN)[37-39]。

由表9可见，即使在恶意攻击、标签噪声与非独立同分布同时存在的情况下，本文方法得到的贡献分布仍与质量分数保持较强的一致性，说明所提出的贡献评价机制具有复杂场景的鲁棒性。

表9　攻击条件下的贡献评价结果

攻击	PCC	Spearman	单调性比例
SF[37]	0.8733	0.7546	0.9142
SA[38]	0.9284	0.8016	0.9371
GN[39]	0.9261	0.7845	0.9257

结合图6可以看出，攻击条件下各类客户端的平均优化权重与贡献值仍保持合理排序。这说明本文方法能够在复杂扰动下识别有害的恶意更新。不过，与无攻击场景相比，图中不同噪声类别之间的区分度有所下降。造成这种现象的原因在于攻击条件下的贡献信号不仅由客户端自身数据质量决定，还受到恶意更新干扰、随机参与组合以及数据非独立同分布的共同影响。首先，恶意客户端会显著扰动全局模型更新方向，使服务器在代理集上观测到的优化信号更偏向于区分明显有害与非明显有害的客户端，而对良性噪声客户端之间的细粒度差异变得不再同样敏感。其次，在随机选取机制下，当某一轮存在多个恶意客户端时，良性但含噪的客户端可能因为相对恶意更新危害更小而表现出接近的边际作用，从而压缩了良性类别之间的差异。

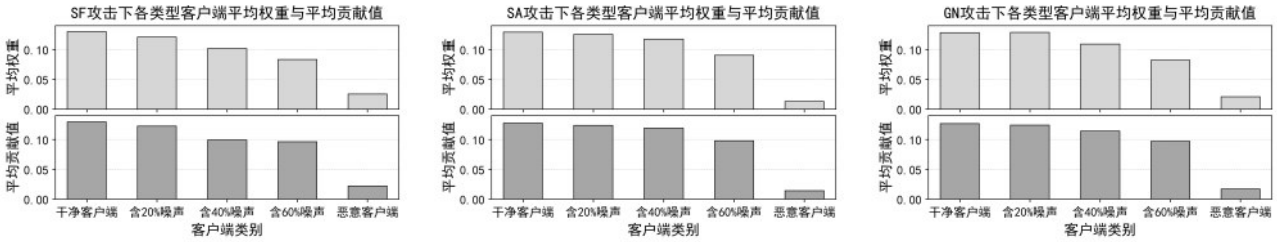

图6　各类客户端的平均优化权重与平均贡献值

图7表明本文方法在三种攻击下均能明显优于受攻击的FedAvg，说明服务器侧权重优化能够在训练阶段有效削弱恶意更新对全局模型的破坏，并在一定程度上保留高质量客户端的正向作用。实验结果说明，本文方法虽然并非专门的拜占庭防御机制，但在典型模型投毒攻击下除了贡献评价鲁棒性之外，本文方案还具备一定的训练鲁棒性。

5.5　方法兼容性与适应性

由图8可知，在存在噪声与Dirichletα = 0.5的复杂环境下，本文方法在三种模态数据集上采用不同聚合算法均能够带来精度提升。相较于FedAvg[6]、FedProx[40]和FedAvgM[41]，加入本文方法后全局模型均有更快的上升趋势和更高的最终精度。这表明，本文方法能够作为一种通用的优化机制，具有较好的聚合算法兼容性和场景适应性。

5.6　参数敏感性分析

表10表明，本文方法对验收集大小变化具有较强稳定性。验收集大小在100－2000的范围内，本文方法的指标始终保持在较高水平，说明方案在验收集大小变化的情况下总体稳定。

验收集偏移对评价效果的影响实验结果如表11所示。当验收集从全类别降至9类时，PCC由0.9340降至0.8280，方法仍保持较高鲁棒性；降至7类和5类时，PCC进一步分别降至0.7505和0.7460，但下降幅度趋缓。仅保留3类时，PCC、Spearman和单调性比例分别降至0.5421、0.5469和0.8114，贡献评价显著退化。说明贡献评价的准确性依赖于验收集对真实任务分布的代表性。

5.7　可解释性说明

表12展示了随机抽取的案例。以客户端3为例，此时位于第24轮，该轮次属于全局模型性能稳定提升的轮次。客户端3第4次被选中，该客户端的历史平滑权重为0.1197，倾向于优质客户端。本轮优化后当轮权重提升至0.1153，表现为正证据，同时其单轮边际作用也表现为正证据。这说明该客户端在当前轮中既获得了服务器侧增权，也对当前轮全局任务产生了正向边际作用，因此其主观逻辑值在同轮参与客户端中位列靠前。其不确定度为0.9187，贡献值向历史先验收缩，最终映射得到当轮贡献值0.1294。

同理其他客户端的贡献结果也可依据上述例子进行解释。

需注意的是，不确定度并非表示当轮观测不可信，而是本文主观逻辑框架内的分配比例机制。不确定度越高，当轮贡献越依赖该客户端的历史平均表现而非瞬时证据。这一设计的核心目的在于避免联邦学习中因随机采样而偶发的局部噪声点过度影响贡献评价。贡献值是跨轮累积的结果，单轮波动由不确定度缓冲，长期收敛方向由持续证据驱动。先验质量常数决定了单轮证据与历史先验之间的分

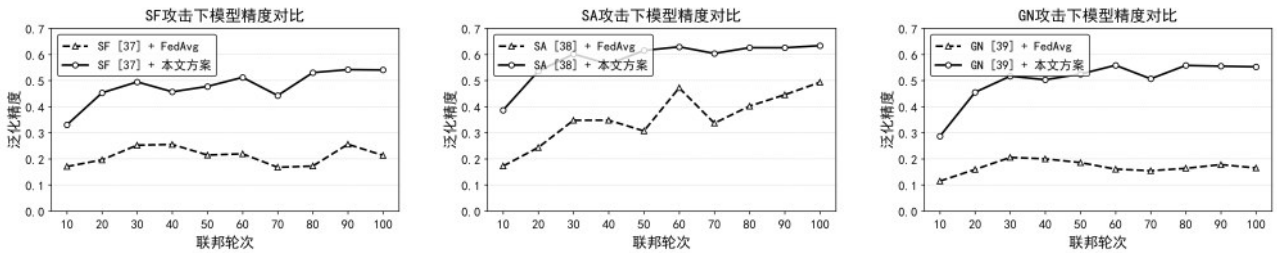

图7　恶意攻击下模型精度对比

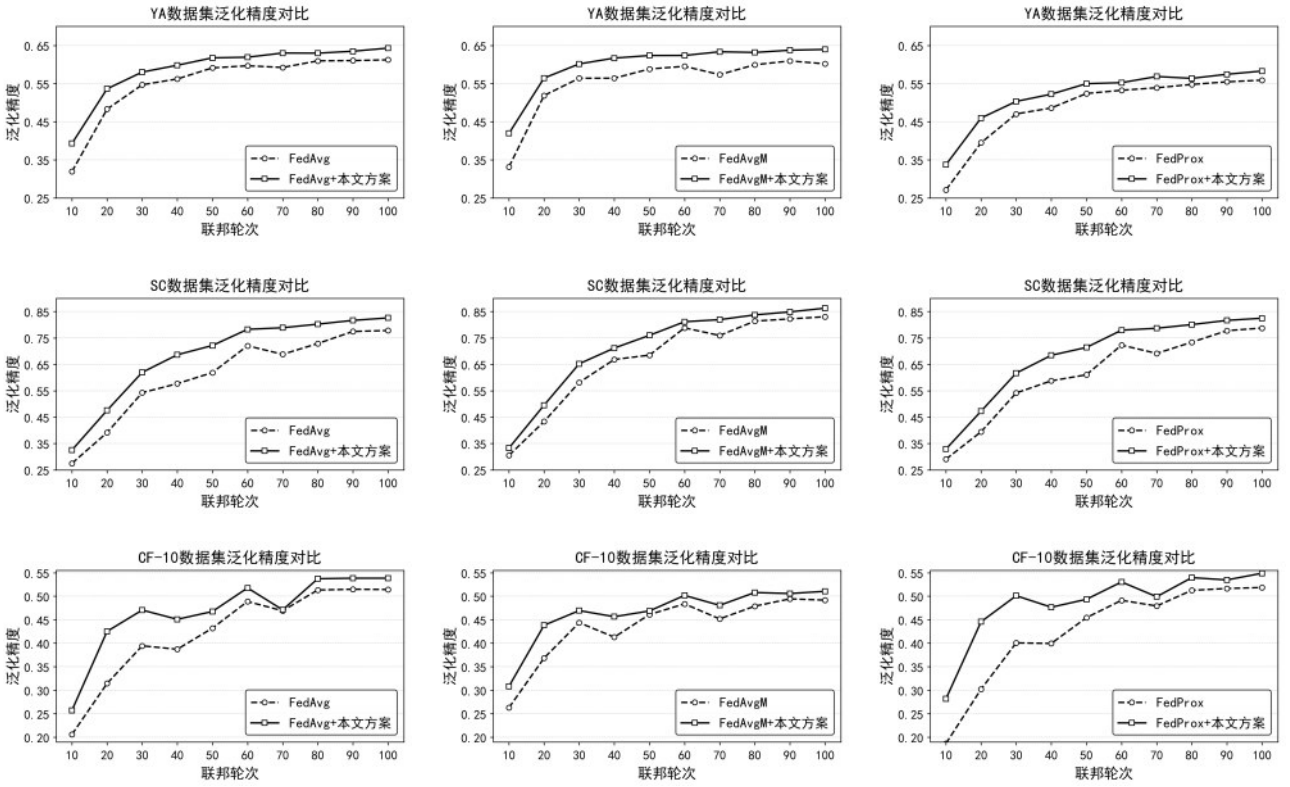

图8　精度优化作用

表10　验收集大小对评价效果的影响

验收集大小	PCC	Spearman	单调性比例	泛化测试集精度
100	0.9429	0.9300	0.9828	0.6240
200	0.9575	0.9451	0.9942	0.6394
300	0.9371	0.9255	0.9787	0.6228
500	0.9548	0.9390	0.9902	0.6208
1000	0.9340	0.9275	0.9812	0.6280
1500	0.9639	0.9531	0.9983	0.6448
2000	0.9621	0.9456	0.9934	0.6144

表11　验收集偏移对评价效果的影响

类别数	PCC	Spearman	单调性比例	泛化测试集精度
全类别	0.9340	0.9275	0.9812	0.6280
9	0.8280	0.8511	0.9330	0.6310
7	0.7505	0.7812	0.9045	0.6383
5	0.7460	0.7666	0.9012	0.6236
3	0.5421	0.5469	0.8114	0.6275

配权重比。

6　结束语

本文面向隐私保护约束下的数据交易场景，提出了一种联邦学习贡献评价机制。所提方法将质量感知的客户端权重优化与主观逻辑映射相结合，能够在不访问原始数据的条件下对客户端有效贡献进行评估，并为数据交易中的数据定价、收益分配与公平激励提供方法支撑。实验结果表明，本文方法在贡献评价准确性、效率和鲁棒性方面表现良好，同时具备动态性与一定的可解释性。

表 12　　主观逻辑映射的贡献评价可解释说明

id	类型	FL 轮次	被选中轮次	历史权重	本轮优化后权重	正证据		负证据		不确定度	**当轮贡献值**
						P_1	P_2	N_1	N_2	U	
3	clean	24	4	0.1197	0.1153	0.0153	0.0377	0.0000	0.0000	0.9187	**0.1292**
	解释：p1=0.0153，n1=0：权重变化为正证据；p2=0.0377，n2=0：边际贡献为正证据；当前 FL 第 24 轮、该客户端第 4 次被选中，历史平滑权重为 0.1197，且在同轮参与者中主观逻辑值位列第 2，当轮贡献被映射为 0.1292。										
24	noise20	55	9	0.0855	0.0823	0.0000	0.0219	0.0177	0.0000	0.9381	**0.0963**
	解释：p1=0，n1=0.0177：出现降低权重负证据；p2=0.0219，n2=0：边际贡献为正证据；结合当前 FL 第 55 轮、该客户端第 9 次被选中，历史平滑权重为 0.0855，且同轮参与者中主观逻辑值位列第 5，当轮贡献被映射为 0.0963。										
33	noise40	33	5	0.1003	0.1018	0.0018	0.0000	0.0000	0.0099	0.9808	**0.0746**
	解释：p1=0.0018，n1=0：出现提升权重正证据；p2=0，n2=0.0165：边际贡献为负证据；结合当前 FL 第 33 轮、该客户端第 5 次被选中，历史平滑权重为 0.1003，且同轮参与者中主观逻辑值位列第 7，当轮贡献被映射为 0.0746。										
41	noise60	27	6	0.0846	0.0921	0.0000	0.0000	0.0079	0.0405	0.9254	**0.0610**
	解释：p1=0，n1=0.0079：出现降低权重负证据；p2=0，n2=0.0811：边际贡献为负证据；结合当前 FL 第 27 轮、该客户端第 6 次被选中，历史平滑权重为 0.0846，且同轮参与者中主观逻辑值位列第 8，当轮贡献被映射为 0.0610。										

参考文献：

[1] 窦悦,徐舒扬,崔佳佳,等.数据流通利用视角下培育全国一体化数据市场路径研究[J].大数据,2026,12(02):18-31.
DOU Yue, XU Shuyang, CUI Jiajia, et al. Research on the path of cultivating a national integrated data market from the perspective of data circulation and utilization[J]. Big Data Research, 2026, 12(02): 18-31.

[2] 乔夫,王庆德.加快我国数据交易所生态系统建设的现实基础与发展路径研究[J].价格理论与实践,2025,(05):88-94+242. DOI:10.19851/j.cnki.CN11-1010/F.2025.05.098.
QIAO Fu, WANG Qingde. Research on the realistic basis and development path for accelerating the construction of China's data exchange ecosystem[J]. Price: Theory & Practice, 2025(05): 88-94+242. DOI:10.19851/j.cnki. CN11-1010/F.2025.05.098.

[3] Feng Z, Yu S, Zhu Y. Towards personalized privacy preference aware data trading: A contract theory based approach[J]. Computer Networks, 2023, 224: 109637.

[4] 周烨.培育五大可信数据空间加速构建全国一体化数据市场——解读《可信数据空间发展行动计划（2024-2028年）》[J].中国科技产业,2025,(1):8-9.DOI:10.16277/j.cnki.cn11-2502/n.2025.01.010.

[5] Bonawitz K, Kairouz P, McMahan B, et al. Federated learning and privacy: Building privacy-preserving systems for machine learning and data science on decentralized data[J]. Queue, 2021, 19(5): 87-114.

[6] McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Artificial intelligence and statistics. Pmlr, 2017: 1273-1282.

[7] Sun P, Liao G, Chen X, et al. A socially optimal data marketplace with differentially private federated learning[J]. IEEE/ACM Transactions on Networking, 2024, 32(3): 2221-2236.

[8] Chai D, Wang L, Yang L, et al. A survey for federated learning evaluations: Goals and measures[J]. IEEE transactions on knowledge and data engineering, 2024, 36(10): 5007-5024.

[9] Ma S, Cao Y, Xiong L. Transparent contribution evaluation for secure federated learning on blockchain[C]//2021 IEEE 37th international conference on data engineering workshops (ICDEW). IEEE, 2021: 88-91.

[10] Chen Y, Li K, Li G, et al. Contributions Estimation in Federated Learning: A Comprehensive Experimental Evaluation[J]. Proc. VLDB Endow., 2024, 17(8): 2077-2090.

[11] 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.

[12] Jiang Y, Zhang W, Chen Y. Data quality detection mechanism against label flipping attacks in federated learning[J]. IEEE Transactions on Information Forensics and Security, 2023, 18: 1625-1637.

[13] 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). IEEE, 2020: 278-288.

[14] 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.

[15] Lyu L, Xu X, Wang Q, et al. Collaborative fairness in federated learning[M]//Federated Learning: Privacy and Incentive. Cham: Springer International Publishing, 2020: 189-204.

[16] 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). IEEE, 2021: 3060-3064.

[17] 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). IEEE, 2022: 2440-2453.

[18] Liu Y, Chang S, Liu Y, et al. FairFed: Improving fairness and efficiency of contribution evaluation in federated learning via cooperative Shapley value[C]//IEEE INFOCOM 2024-IEEE Conference on Computer Communications. IEEE, 2024: 621-630.

[19] Lei K, Ren X, Yang S, et al. Feddsv: Shapley value-based contribution estimation in federated learning with dynamic participation[J]. IEEE Transactions on Mobile Computing, 2025.

[20] Wu H, Zhang L, Li S, et al. CoAst: Validation-Free Contribution Assessment for Federated Learning based on Cross-Round Valuation[C]// Proceedings of the 32nd ACM International Conference on Multimedia. 2024: 1839-1847.

[21] Jiang M, Roth H R, Li W, et al. Fair federated medical image segmentation via client contribution estimation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 16302-16311.

[22] Guo J, Li J, Liu Z, et al. LCEFL: A lightweight contribution evaluation approach for federated learning[J]. IEEE Transactions on Mobile Computing, 2025.

[23] Yang Y, Xiao P, Ma S, et al. First-order federated bilevel learning[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2025, 39(21): 22029-22037.

[24] Xie Y, Zhang W, Pi R, et al. Optimizing server-side aggregation for robust federated learning via subspace training[J]. arXiv preprint arXiv: 2211.05554, 2022, 2(9).

[25] Huang W, Shi Z, Ye M, et al. Self-driven entropy aggregation for byzantine-robust heterogeneous federated learning[C]//Forty-first International Conference on Machine Learning. 2024.

[26] Li Z, Lin T, Shang X, et al. Revisiting weighted aggregation in federated learning with neural networks[C]//International Conference on Machine Learning. PMLR, 2023: 19767-19788.

[27] Parsa J, Daghestani A H, Teixeira A M H, et al. Byzantine-Robust Federated Learning with Learnable Aggregation Weights[J]. arXiv preprint arXiv:2511.03529, 2025.

[28] Nourmohammadi R, Zhang K, Talhi C. BlockFed: A novel federated learning framework based on hierarchical weighted aggregation[C]// Proceedings of the 19th ACM International Conference on Distributed and Event-based Systems. 2025: 134-145.

[29] Shi C, Zhao H, Zhang B, et al. FedAWA: adaptive optimization of aggregation weights in federated learning using client vectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2025: 30651-30660.

[30] Kairouz P, McMahan H B. Advances and open problems in federated learning[J]. Foundations and trends in machine learning, 2021, 14(1-2): 1-210.

[31] 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.

[32] Jsang A. Subjective Logic: A formalism for reasoning under uncertainty[M]. Springer Publishing Company, Incorporated, 2018.

[33] Rashid M M, Xiang Y, Uddin M P, et al. Trustworthy and fair federated learning via reputation-based consensus and adaptive incentives[J]. IEEE Transactions on Information Forensics and Security, 2025.

[34] Roy P, Sarker S, Razzaque M A, et al. Distributed task allocation in mobile device cloud exploiting federated learning and subjective logic [J]. Journal of Systems Architecture, 2021, 113: 101972.

[35] Song T, Tong Y, Wei S. Profit allocation for federated learning[C]// 2019 IEEE International Conference on Big Data (Big Data). IEEE, 2019: 2577-2586.

[36] Wang G, Dang C X, Zhou Z. Measure contribution of participants in federated learning[C]//2019 IEEE international conference on big data (Big Data). IEEE, 2019: 2597-2604.

[37] Damaskinos G, Guerraoui R, Patra R, et al. Asynchronous Byzantine machine learning (the case of SGD)[C]//International Conference on Machine Learning. PMLR, 2018: 1145-1154.

[38] Bagdasaryan E, Veit A, Hua Y, et al. How to backdoor federated learning[C]//International conference on artificial intelligence and statistics. PMLR, 2020: 2938-2948.

[39] Fang M, Cao X, Jia J, et al. Local model poisoning attacks to {Byzantine-Robust} federated learning[C]//29th USENIX security symposium (USENIX Security 20). 2020: 1605-1622.

[40] 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.

[41] Hsu T M H, Qi H, Brown M. Measuring the effects of non-identical data distribution for federated visual classification[J]. arXiv preprint arXiv:1909.06335, 2019.

[42] Wang J, Zhang L, Li A, et al. Efficient participant contribution evaluation for horizontal and vertical federated learning[C]//2022 IEEE 38th International Conference on Data Engineering (ICDE). IEEE, 2022: 911-923.

田有亮（1982- ），男，博士，贵州大学教授、博士生导师，主要研究方向为博弈论、密码学与安全协议、大数据隐私保护等。

支睿（2000- ），男，贵州大学硕士生，主要研究方向为隐私保护、分布式学习、激励机制、合作博弈等。

王帅（2000- ），男，贵州大学博士生，主要研究方向为隐私保护机器学习、分布式学习和安全聚合。

金昆龙（2000– ），男，贵州大学硕士生，主要研究方向为隐私保护、联邦学习、后门攻击。