

面向异构数据的具有隐私保护性能的公平联邦学习

吴炬¹, 苗银宾¹, 张炜斌¹, 刘淑英^{2,3}, 马建峰¹

(1. 西安电子科技大学网络与信息安全学院, 陕西 西安 710126;

2. 西北大学计算机学院, 陕西 西安 710127;

3. 咸阳师范学院计算机学院, 陕西 咸阳 712000)

摘要: 联邦学习在非独立同分布数据场景下易因梯度冲突导致公平性失衡, 且模型参数传输存在隐私泄露风险。现有公平性联邦学习算法虽有效缓解了梯度冲突, 但复杂异质数据下局部与全局模型的一致性缺乏针对性优化, 导致局部训练偏离全局最优, 进而削弱公平性并限制整体模型性能。针对此问题, 提出FedALF及LDP-FedALF算法。使用公平驱动目标和分层优化缓解梯度冲突; 引入交替方向乘子法进行优化, 在客户端局部目标函数中设计对偶变量项与二次惩罚项, 构建局部与全局模型一致性约束, 进一步提升公平性与模型性能; 同时在模型传输中加入自适应本地差分隐私技术保护数据隐私。通过大量实验验证得出, FedALF相较于现有公平联邦学习算法, 测试准确率平均提升10.82%, 公平性平均提升19.80%; LDP-FedALF相较于现有公平联邦学习算法, 测试准确率平均提升7.53%, 公平性平均提升15.36%。

关键词: 联邦学习; 公平性优化; 交替方向乘子法; 隐私保护

中图分类号: TP309

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.AQ26019

Fair federated learning with privacy protection for heterogeneous data

Wu Ju¹, Miao Yinbin¹, Zhang Weibin¹, Liu Shuying^{2,3}, Ma Jianfeng¹

1. School of Cybersecurity, Xidian University, Xi'an 710126, China

2. School of Computer Science, Northwest University, Xi'an 710127, China

3. School of Computer Science, Xianyang Normal University, Xianyang 712000, China

Abstract: Federated learning suffers from fairness imbalance due to gradient conflicts in non-independent and identically distributed (Non-IID) data scenarios, with model parameter transmission risking privacy leakage. Existing fair federated learning algorithms mitigate gradient conflicts effectively but lack targeted optimization for local-global model consistency under complex heterogeneous data, leading to deviated local training from global optima, impaired fairness and limited overall performance. To address this, the FedALF algorithm and LDP-FedALF algorithm were proposed: it alleviates gradient conflicts via fairness-driven objectives and hierarchical optimization; adopts ADMM for optimization, designing dual variables and quadratic penalty terms in clients' local objective functions to enforce local-global model consistency constraints, further enhancing fairness and performance; and integrates adaptive local differential privacy into model transmission for privacy protection. Through a large number of experiments, FedALF achieved an average improvement of 10.82% in test accuracy and 19.80% in fairness per-

收稿日期: 2026-01-21; 修回日期: 2026-04-07

通信作者: 刘淑英, lsynate@163.com

基金项目: 大学生创新创业训练计划 (No.S202510722063)

Foundation Item: College Students' Innovation and Entrepreneurship Training Program (No.S202510722063)

formance compared to existing fair federated learning algorithms. LDP-FedALF achieved an average improvement of 7.53% in test accuracy and 15.36% in fairness performance compared to existing fair federated learning algorithms.

Key words: federated learning, fairness optimization, alternating direction method of multipliers, privacy-preserving

0 引言

联邦学习^[1-2]是隐私计算领域的核心技术范式,其核心理念是不传输原始数据,实现多客户端协同训练全局模型。这种“数据不动模型动”的分布式协作模式,成功平衡了数据隐私保护与模型性能优化的需求,已经在自然语言处理^[3]、目标检测^[4]、人脸识别^[5]等众多领域展现出巨大的应用价值。然而联邦学习在实际部署中面临显著异构性瓶颈,主要体现在两方面^[6]:一是系统异构,即参与训练的客户端在计算能力、网络带宽、存储资源等系统资源上存在显著差异;二是数据异构,即参与训练的客户端数据非独立同分布(non-independent and identically distributed, Non-IID)。异构性催生了联邦学习系统公平性的多元需求^[7],包括贡献公平、参与机会公平、模型性能公平。

本文聚焦模型性能公平,即客户端之间测试精度分布集中且波动较小^[8]。该公平性内涵旨在避免数据分布差异导致部分客户端获得显著更差的模型服务。这反映了联邦学习作为协作式学习范式的核心诉求——各参与方应获得相对一致的模型质量,而非仅追求平均性能最优。目前研究者对联邦学习公平性的研究已经取得了一些进展。文献[9]提出无偏联邦学习框架,其优化目标为适配客户端数据分布的任意混合分布,而非仅拟合多数客户端,从而实现模型对不同客户端的性能均衡。但是,文献[9]未显式处理梯度冲突,针对这个不足,文献[10]提出联邦公平平均(federated fair averaging, FedFV)算法,通过余弦相似度检测梯度冲突,调整梯度方向与幅度以消除冲突,平衡公平性、准确率与效率。为了进一步从数学严谨性上保证冲突消解,文献[11]将联邦学习建模为多目标优化问题,提出联邦多梯度下降+算法(federated multi-gradient descent algorithm plus, FedMGDA+)。该算法实现简单、

超参数调节少,可收敛至帕累托平稳解且不牺牲任何用户性能。在此基础上,文献[12]进一步提出联邦多方向公平指导(federated multi-directional fair guidance, FedMDFG)算法,通过多梯度下降与公平指导目标计算无冲突的公平下降方向,并设计低通信成本的步长线搜索策略,同时兼顾公平性与收敛性。但上述方案仅聚焦于模型级梯度冲突,未能关注深度神经网络中浅层负责局部特征提取、深层聚焦全局特征融合的层级功能差异性。这种差异性会导致客户端在不同层级的梯度更新方向出现分歧,进而引发层级梯度冲突^[13-15],影响局部与全局模型的层级协同效果。为了应对此问题,文献[16]提出了分层优化公平联邦学习(layer-wise fair federated learning, FedLF)算法,通过静态公平驱动目标和分层优化缓解了模型级和层级梯度冲突。

目前的公平性联邦学习主要存在以下两个问题。

(1) 局部与全局模型一致性不足。现有公平联邦学习算法虽通过公平驱动目标^[12, 16]等方法缓解了梯度冲突,但在复杂异构数据场景下,未对局部模型与全局模型的一致性进行针对性优化。

(2) 现有公平性联邦学习算法^[16, 25]在设计过程中大多未充分考虑隐私保护^[17]需求,仅聚焦于公平性与性能的平衡,却忽视了模型参数(或梯度)传输过程中的隐私泄露风险——攻击者可通过梯度反推等技术还原客户端原始数据^[18],严重限制了算法在隐私敏感环境下的应用。

在联邦学习的异构数据场景下,梯度冲突是公平性受损的主要原因。模型级梯度冲突表现为梯度方向对立,聚合更新偏向部分客户端;层级梯度冲突源于神经网络各层功能差异,表现为各客户端在不同层级的梯度更新方向出现分歧。若仅处理模型级冲突而忽略层级梯度冲突,则无法从根本上缓解公平性受损问题。更为关键的是,当缺乏一致性约束时,客户端会过度适应本地数

据分布,使本地模型参数空间逐渐偏离全局最优区域。这种偏离使得后续轮次中客户端梯度方向更加分散,从而加剧模型级与层级梯度冲突,甚至出现“越训练越不公平”的现象。因此本文选择引入一致性约束来缓解这个现象。

需要指出的是,传统联邦学习中关于局部与全局模型一致性的研究主要聚焦于模型参数的一致性,以提升收敛速度和缓解Non-IID带来的精度下降,其一致性目标不涉及公平性约束。而本文的一致性是在公平性驱动目标下的更新方向对齐:局部模型更新不仅需要与全局参数保持一致,还要避免偏离公平优化轨迹,防止客户端沿性能分化方向偏移。这使得一致性约束成为与公平更新方向协同优化的动态过程,这是传统方法面临的新挑战。

针对现有公平联邦学习算法局部与全局模型一致性不足和缺乏隐私保护两大问题,本文提出一种基于交替方向乘子法优化的隐私保护公平联邦学习(LDP-FedALF)算法。本文的贡献如下。

(1) 提出了基于交替方向乘子法(alternating direction method of multipliers, ADMM)优化的隐私保护分层公平联邦学习框架,解决了传统公平性联邦学习未兼顾隐私保护和模型性能优化的问题。

(2) 引入ADMM优化模块,将对偶变量项与二次惩罚项融入客户端局部目标函数,有效抑制局部训练偏离全局最优方向的问题,基于现有公平优化逻辑,在进一步提升模型公平性的同时提升全局模型整体性能。相比于传统仅依靠近端正则化的一致性方法,ADMM利用对偶分解机制将复杂的全局约束解耦为可并行求解的局部子问题,不仅显著降低了高维参数下的计算复杂度,更在数学严谨性上保障了分布式场景下局部更新与全局公平目标的高度一致。

(3) 在客户端局部模型参数上传前,采用自适应本地差分隐私(local differential privacy, LDP)机制,通过引入更少的噪声实现隐私保护,在保证模型精度的同时提升算法在隐私敏感环境下的可行性。

(4) 在Fashion-MNIST、CIFAR-10及CIFAR-100等公开数据集的不同异质数据场景下,对本文算法进行系统性测试。实验结果表明,该算法在复杂异质场景下的公平性指标与平均测试准确

率均优于现有公平联邦学习算法。

1 相关工作

作为联邦学习的基础算法,联邦平均(federated averaging, FedAvg)算法^[1]通过客户端本地随机梯度下降(stochastic gradient descent, SGD)与服务器参数加权平均实现去中心化隐私保护训练,但其依赖独立同分布(independent and identically distributed, IID)假设,在实际Non-IID数据场景下易引发梯度抵消、收敛缓慢及个性化不足等问题。为了应对这一不足,文献[6]提出的近端联邦优化(federated proximal, FedProx)算法,在FedAvg算法基础上引入近端正则项,通过约束客户端局部模型与全局模型的参数偏差,有效缓解传统联邦学习收敛缓慢、泛化能力下降的问题。但其未充分考虑客户端数据分布的几何差异。针对此不足,文献[19]提出了一种新颖的层次化Sinkhorn距离聚类联邦学习算法,以最优传输代价衡量数据分布,通过层次化聚类策略动态调整聚合权重,提升高异构场景下的模型准确率与鲁棒性。为了更精准地过滤无效更新,文献[20]提出了一种融合自适应聚合、正则化与元学习优化的联邦学习(federated learning with adaptive aggregation, regularization and meta-learning optimization, FARMO)算法,有效地抑制低质量更新并增强模型自适应能力。随着数据异构程度的加剧,单纯的聚合优化已难以维持模型性能,研究者开始引入知识蒸馏与强化学习等先进机制。文献[21]提出了基于知识迁移的联邦学习(federated learning via knowledge transfer, FedKT)算法,通过模型分割蒸馏与一致性约束,缓解Non-IID导致的性能退化。文献[22]提出了基于强化学习与知识蒸馏的非独立同分布数据联邦学习(federated learning on Non-IID data via reinforcement learning and knowledge distillation, FedAgent)算法,结合强化学习客户端选择与知识蒸馏,解决Non-IID环境下精度退化、收敛缓慢等问题。此外,文献[33]提出了基于交替方向乘子法的联邦优化算法(federated learning algorithm based on alternating direction method of multipliers, FedADMM)框架,通过在局部客户端子问题中引入对偶变量,有效提升了联邦学习在系统

和数据异质环境下的收敛稳定性。

然而, 这些方法并未考虑联邦学习过程中的公平性问题, 即最终模型在客户端上的表现效果差异巨大, 导致边缘客户端参与积极性下降, 进而影响模型整体性能。因此, 研究者们进行了一系列公平性算法的研究。文献[23]提出分位数联邦平均 (quantized federated averaging, qFedAvg) 算法, 针对联邦学习中数据异构与资源异构引发的客户端性能失衡问题, 通过贡献感知的权重分配机制, 解决了数据与资源异构引发的客户端性能失衡问题, 有效改善了异构场景下客户端间的性能差异。但这种强制性的均衡往往难以保障每一个参与方都能获得最优的模型体验。为了实现一种更具包容性且能兼顾个体差异的公平分配, 文献[8]提出了 Ditto 个性化联邦学习算法, 通过局部个性化模型训练和全局共识约束, 解决了传统联邦学习在 Non-IID 场景下性能失衡和客户端之间的公平性问题。在公平与隐私的协同方面, 文献[24]提出了公平与差分隐私联邦学习 (federated learning with fairness and differential privacy, FedFDP) 算法, 通过公平感知梯度裁剪、损失自适应裁剪及最优公平性参数, 并结合 Rényi 差分隐私进行隐私分析, 解决了联邦学习中公平性、差分隐私保护与模型性能难以协同优化的问题, 实现了三者的高效平衡。为了克服静态策略在复杂环境下的局限性, 文献[25]提出了融合动态学习率调整、策略性客户端选择、局部模型自适应与伪标签技术的公平动态个性化联邦学习 (fair dynamic personalized federated learning with strategic client selection, FairDPFL-SCS) 框架, 有效解决了联邦学习中 Non-IID 数据导致的模型个性化适配不足、客户端间公平性失衡及训练效

率低下下的问题。

然而, 目前的公平性研究大多只关注模型级的梯度冲突, 未关注模型的层级冲突。FedLF 算法通过分层策略解决了这个问题, 然而其忽略了 Non-IID 设置带来的全局模型与局部模型的一致性问题, 导致模型在高数据异质环境下稳定性和鲁棒性不足。因此, 本文将 ADMM 算法引入 FedLF 算法来缓解 Non-IID 数据带来的影响, 同时加入自适应 LDP 机制, 进一步提升现有公平性联邦学习的模型整体性能、公平性和安全性。现有公平联邦学习算法的比较见表 1, FedALF 为基于交替方向乘子法优化的公平联邦学习算法。

2 预备知识

2.1 联邦学习

联邦学习的目标是求解全局模型 ω , 使所有客户端的加权平均损失最小化, 公式化描述形式如下:

$$\begin{cases} \min_{\omega \in \mathbf{R}^d} F(\omega) = \sum_{i=1}^m \frac{n_i}{n} f_i(\omega) \\ f_i(\omega) = \frac{1}{n_i} \sum_{\xi_k \in D_i} \mathcal{L}(\omega, \xi_k) \end{cases} \quad (1)$$

其中, D_i 为第 i 个客户端的本地数据集, ξ_k 为 D_i 中的第 k 个样本, n_i 为 D_i 的样本量, $n = \sum_{i=1}^m n_i$ 为所有客户端的样本总量, $\mathcal{L}(\omega, \xi_k)$ 为模型训练损失函数, ω 为全局模型参数, m 为参与客户端的总数, $\mathbf{R}^d$ 为 d 维实向量空间, $f_i(\omega)$ 为客户端 i 的本地损失, $F(\omega)$ 为客户端损失加权求和。

2.2 ADMM

ADMM^[26-27]是由 Douglas-Rachford 分裂算法与 Bregman 迭代算法融合衍生的经典分布式优化技术, 其核心优势在于能将高复杂度的带约束优

表 1 现有公平联邦学习算法的比较

方案	核心思想	是否解决层级冲突	是否进行隐私保护
文献[12]	公平目标约束结合多梯度下降	否	否
文献[23]	公平感知权重分配	否	否
文献[24]	公平感知梯度裁剪、差分隐私保护	否	是
文献[25]	动态学习率调整、客户端选择	否	否
文献[16]	公平目标驱动、分层优化	是	否
FedALF	公平目标驱动、分层优化、ADMM 优化	是	否
LDP-FedALF	公平目标驱动、分层优化、ADMM 优化、自适应 LDP	是	是

化问题拆解为多个可独立求解的子问题，在降低求解难度的同时适配分布式计算与并行处理场景。因此，ADMM在联邦学习、信号处理等需要协同优化的领域被广泛应用。

首先针对带线性约束的凸优化问题构建统一数学模型。假设待求解的优化目标是最小化由两个凸函数 $f(\cdot)$ 与 $g(\cdot)$ 构成的目标函数，同时满足线性等式约束 $\mathbf{Ax} + \mathbf{Bz} = \mathbf{c}$ ，其数学表达式如下：

$$\begin{cases} \min f(\mathbf{x}) + g(\mathbf{z}) \\ \text{s.t. } \mathbf{Ax} + \mathbf{Bz} = \mathbf{c} \end{cases} \quad (2)$$

其中， $\mathbf{x} \in \mathbf{R}^n$ 与 $\mathbf{z} \in \mathbf{R}^m$ 为待优化向量，分别对应分布式场景下不同节点的局部参数； $\mathbf{A} \in \mathbf{R}^{p \times n}$ 、 $\mathbf{B} \in \mathbf{R}^{p \times m}$ 为已知系数矩阵， $\mathbf{c} \in \mathbf{R}^p$ 为约束目标向量。为有效处理上述约束优化问题，ADMM引入增广拉格朗日函数，将约束条件融入目标函数，其定义如下：

$$\begin{aligned} \mathcal{L}(\mathbf{x}, \mathbf{y}, \mathbf{z}) &= f(\mathbf{x}) + g(\mathbf{z}) + \mathbf{y}^T \\ &(\mathbf{Ax} + \mathbf{Bz} - \mathbf{c}) + \frac{\rho}{2} \|\mathbf{Ax} + \mathbf{Bz} - \mathbf{c}\|^2 \end{aligned} \quad (3)$$

其中， $\mathcal{L}(\cdot)$ 即为增广拉格朗日损失函数， $\mathbf{y} \in \mathbf{R}^p$ 是拉格朗日乘子向量，用于衡量约束条件的违反程度。二次惩罚系数 ρ 用于平衡约束满足度与目标函数优化速度。下面以第 k 轮为例，首先固定当前轮次的 $\mathbf{z}^k$ 和 $\mathbf{y}^k$ ，求解 $\mathbf{x}$ 的局部最优解（记为 $\mathbf{x}^{k+1}$ ）。此时 $\mathbf{x}$ 的优化子问题的数学表达式为：

$$\mathbf{x}^{k+1} = \arg \min_{\mathbf{x}} \left(f(\mathbf{x}) + \frac{\rho}{2} \left\| \mathbf{Ax} + \mathbf{Bz}^k - \mathbf{c} + \frac{\mathbf{y}^k}{\rho} \right\|^2 \right) \quad (4)$$

在得到 $\mathbf{x}^{k+1}$ 后，下一步固定 $\mathbf{x}^{k+1}$ 与 $\mathbf{y}^k$ ，对 $\mathbf{z}$ 执行类似的最小化操作，求解 $\mathbf{z}$ 的局部最优解（记为 $\mathbf{z}^{k+1}$ ）。该过程同样仅依赖 $g(\cdot)$ 与已更新的 $\mathbf{x}^{k+1}$ ，对应的数学表达式为：

$$\mathbf{z}^{k+1} = \arg \min_{\mathbf{z}} \left(g(\mathbf{z}) + \frac{\rho}{2} \left\| \mathbf{Ax}^{k+1} + \mathbf{Bz} - \mathbf{c} + \frac{\mathbf{y}^k}{\rho} \right\|^2 \right) \quad (5)$$

最后，基于 $\mathbf{x}^{k+1}$ 和 $\mathbf{z}^{k+1}$ 计算新一轮 $\mathbf{y}^{k+1}$ ：

$$\mathbf{y}^{k+1} = \mathbf{y}^k + \rho(\mathbf{Ax}^{k+1} + \mathbf{Bz}^{k+1} - \mathbf{c}) \quad (6)$$

上述过程完成了一次ADMM算法的迭代，重复执行此过程进行后续更新。

2.3 自适应LDP

差分隐私（differential privacy, DP）^[28]是一

种通过扰动原始数据来保护数据隐私的标准范式，这样任何攻击者都无法从获得的数据中推断出原始隐私信息。自适应LDP^[29]的定义如下。

假设 $\mathcal{M}$ 为一个随机扰动机制，对于任意一对输入 x 和 x' ，以及 $\mathcal{M}$ 的任意一个输出 Y ，如果 $\mathcal{M}$ 满足 ϵ -LDP，则必须符合以下不等式

$$\Pr[\mathcal{M}(x) = Y] \leq e^\epsilon \cdot \Pr[\mathcal{M}(x') = Y] \quad (7)$$

其中， ϵ 是 $\mathcal{M}$ 的隐私预算， ϵ 的值越小，表示隐私保护程度越强，但也意味着数据可用性越低^[30]。

2.4 主要符号表

本文所用的主要符号见表2。

表2 主要符号

符号	描述
m	客户端数量
ρ	惩罚系数
ϵ	隐私预算
C^t	第 t 轮参与训练的客户端集合
y_i^t	客户端 i 在第 t 轮的对偶变量
ω_i^t	客户端 i 在第 t 轮的本地模型
ω_i^{t*}	客户端 i 在第 t 轮加噪的本地模型
g_p^t	第 t 轮公平驱动目标的梯度

3 问题描述

3.1 系统模型

方案采用“客户端-服务器”二级架构，核心由参与训练的客户端与负责全局协调的服务器两类实体构成，二者通过传递经过自适应LDP保护的模型参数实现协同训练，全程不传输原始数据以保障隐私安全。

服务器：中心服务器是服务提供者，负责协调客户端之间的间接数据交换和进行全局模型的聚合。

客户端：客户端是一组申请协同服务的请求者，负责进行局部模型训练并对训练结果完成自适应LDP保护。

系统模型如图1所示。

(1) 服务器按预设策略采样客户端，向选中客户端下发当前全局模型参数。

(2) 客户端以本地数据为基础，结合ADMM约束优化局部模型，生成局部更新并通过注入噪声的方式实现自适应LDP保护。

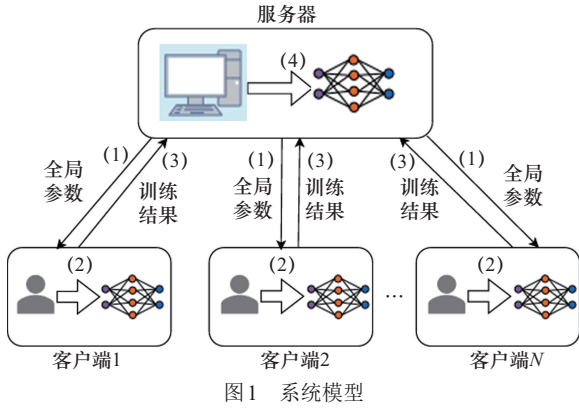


图1 系统模型

(3) 客户端将添加噪声的局部模型参数、对偶变量及损失信息上传至服务器。

(4) 服务器融合局部更新，计算公平且无梯度冲突的全局更新方向，更新全局模型。

重复上述步骤，服务器将新的全局模型参数广播至客户端，客户端同步更新后进入下一轮迭代，直至模型收敛。

3.2 威胁模型

在本文的系统中，服务器和客户端是诚实但好奇的，具体来说，所有实体都诚实地遵循最初设置的学习协议，但服务器会通过客户端的本地模型推断客户端的隐私信息，客户端也可能通过全局模型尝试推断其他客户端的隐私信息。此外，客户端的数据统计异构性强，上述实体可能造成如下潜在威胁。

(1) 模型精度和公平性下降：客户端的数据集呈现 Non-IID 分布，不仅会使模型的公平性受损，还会导致全局模型的精度降低。

(2) 数据隐私泄露：中央服务器是诚实但好奇的，服务器会通过客户端的本地模型推断客户端的隐私信息。

3.3 问题定义

本文考虑由 1 个中央服务器与 m 个客户端组成的“服务器-客户端”二级联邦学习架构，记客户端集合为 $\mathcal{K} = \{1, 2, \dots, m\}$ 。每个客户端 $k \in \mathcal{K}$ 持有私有数据集 $D_k = \{(x_{k,i}, y_{k,i})\}_{i=1}^{n_k}$ ，其中 $n_k = |D_k|$ 为客户端 k 的样本数量。设全局模型参数为 $\omega \in \mathbf{R}^d$ (d 为模型参数维度)，客户端 k 的局部模型参数为 $\omega_k \in \mathbf{R}^d$ ，局部模型与全局模型有相同网络结构。联邦学习的核心流程为迭代优化：服务器第 t 轮选择子集 $C^t \subseteq \mathcal{K}$ 参与训练，下

发当前全局模型 ω^t ；客户端基于 D_k 在 ADMM 优化下完成局部训练，随后使用 LDP 机制对模型加噪，将加噪后的局部模型 $\omega_k^{(t+1^*)}$ 、更新后的对偶变量以及局部损失值上传至服务器；服务器通过公平目标驱动和分层聚合策略更新全局模型 ω^{t+1} ，迭代直至模型收敛。

联邦学习的公平性定义为全局模型在所有参与客户端上的性能均衡性。公平性指标定义为 $A(\omega)$ 和 $\mathbf{1}$ 之间的夹角，令其为 θ ，表示如下：

$$\theta = \arccos \left(\frac{A(\omega) \cdot \mathbf{1}}{\|A(\omega)\| \|\mathbf{1}\|} \right) \quad (8)$$

其中， $A(\omega) = (A_1(\omega), A_2(\omega), \dots, A_n(\omega))$ 为全局模型在各客户端上的准确率构成的向量，其中 $A_n(\omega)$ 是当前模型在客户端 n 的准确率， $\mathbf{1} = (1, 1, \dots, 1)$ 是和 $A(\omega)$ 相同维度的全 1 向量。 $A(\omega)$ 与全 1 向量 $\mathbf{1}$ 之间夹角越小，代表模型在客户端间的性能均衡性越好，公平性越优。该指标与方差、基尼系数等传统公平性指标在本质上一致，均反映了客户端性能的均衡程度。选用本文指标的核心原因在于，其基于余弦相似度衡量准确率向量与理想公平状态的方向契合度，既能规避传统指标对极端值的过度敏感，又能有效识别“少数客户端占优、多数客户端受益不足”的伪公平场景，更精准契合联邦学习“所有参与方均衡受益”的核心需求，同时具备直观的物理意义与严谨的理论支撑，为公平性评估提供更鲁棒、更贴合实际场景的量化依据。

模型级梯度冲突定义为：客户端全局梯度方向对立导致聚合更新方向损害部分客户端利益。如 g_i^t 和 g_j^t 分别是第 t 轮客户端 i 和客户端 j 的梯度，它们满足 $g_i^t \cdot g_j^t < 0$ ，本文认为其为模型级别的冲突，最终更新方向将会与其中一个的梯度方向冲突。

层级梯度冲突定义为：深度神经网络各层功能差异引发客户端层级梯度分歧。如 $g_{i,k}^t$ 和 $g_{j,k}^t$ 分别是第 t 轮客户端 i 和客户端 j 的第 k 层的梯度片段， d_k^t 是更新方向在第 k 层的分量， $d_k^t \cdot g_{i,k}^t > 0$ 且 $d_k^t \cdot g_{j,k}^t < 0$ ，则最终更新方向在第 k 层不利于客户端 i 而有利于客户端 j 。

3.4 设计目标

本文方案的核心设计目标如下。

(1) 公平性：全局模型应在不同客户端之间实现性能均衡，即各客户端的测试准确率分布集中、波动较小，避免数据异构导致部分客户端被系统性忽视。

(2) 隐私安全：在模型参数传输过程中，应抵御来自服务器或其他客户端的隐私推断攻击，在满足严格形式化隐私保障的前提下，确保攻击者无法从上传参数中有效还原客户端的原始数据或敏感属性，满足 ϵ -LDP 隐私保护要求。

(3) 模型效用：在保障公平性与隐私安全的前提下，方案应保持较高的全局模型精度与稳定的收敛性能，避免公平性约束或隐私保护机制引入过多噪声或优化偏差，导致模型实用性下降。

4 方案设计

4.1 技术思路

为解决传统联邦学习的公平性缺陷，首先须明确其本质是多目标优化问题，即最小化由各客户端局部目标函数构成的向量：

$$\min_{\omega} (F_1(\omega), F_2(\omega), \dots, F_m(\omega)) \quad (9)$$

FedMDFG^[12]通过动态目标函数增强多目标优化公平性，核心是依据模型状态调整优化方向，引入公平驱动目标，通过计算目标值分布与理想公平状态的偏移生成反向修正向量，避免客户端性能失衡。但该方法存在明显不足：客户端目标值接近时，该修正机制对微小差异过度放大，反而引入不必要的偏移，导致优化方向偏离真实公平最优解；引入的超参数影响算法表现，导致收敛性难以保证，在实际部署中须反复调参才能获得稳定性能。

为弥补这些缺陷，FedLF 设计了更有效的静态公平驱动目标，核心是通过最小化 $P(\omega) = -\cos(\mathbf{1}, \mathbf{F}(\omega))$ ，提高全 1 向量 $\mathbf{1}$ 与目标向量 $\mathbf{F}(\omega) = (F_1(\omega), \dots, F_m(\omega))$ （由各客户端局部目标函数构成）的余弦相似度，进而缩小客户端间目标函数的差异。因此，FedLF 的优化目标可以扩展为如下多目标优化问题

$$\min_{\omega} (F_1(\omega), F_2(\omega), \dots, F_m(\omega), P(\omega)) \quad (10)$$

相较于 FedMDFG 的动态目标，该静态目标

不需要复杂的动态调整机制，更便于分析收敛性及收敛速率。关键在于，对于原问题的非帕累托平稳解 ω' ，始终存在一个优化方向 $\mathbf{d}'$ ，既能保证所有客户端的梯度与该方向形成公共下降关系，又能降低公平驱动目标，这一特性确保新增公平目标不会影响联邦学习的收敛性。其中，第 t 轮公平驱动目标的梯度 $\mathbf{g}_p^t$ 计算公式如下：

$$\mathbf{g}_p^t = \frac{\mathbf{g}^t}{\|\mathbf{F}(\omega')\|^2} \cdot \left(\frac{\mathbf{F}(\omega')^T \mathbf{1} \cdot \mathbf{F}(\omega')}{\|\mathbf{1}\| \cdot \|\mathbf{F}(\omega')\|} - \frac{\mathbf{1} \cdot \|\mathbf{F}(\omega')\|}{\|\mathbf{1}\|} \right) \quad (11)$$

其中， $\mathbf{g}^t$ 为所有客户端梯度的拼接向量。

FedLF 采用 $\omega^{t+1} = \omega^t + \eta^t \mathbf{d}^t$ 的迭代方式更新模型，其中 $\mathbf{d}^t$ 是第 t 轮的分层公平更新方向，整个方案的核心在于该方向的精准计算与优化。首先明确两类梯度冲突的定义：模型级梯度冲突指客户端 i 和 j 的梯度满足 $\mathbf{g}_i \cdot \mathbf{g}_j < 0$ ，而层级梯度冲突则是客户端在网络层 k 的梯度片段满足 $\mathbf{g}_{i,k} \cdot \mathbf{g}_{j,k} < 0$ （其中 $\mathbf{g}_{i,k}$ 为客户端 i 在层 k 的梯度片段）。现有公平联邦学习算法仅聚焦模型级梯度冲突，无法避免层级更新方向偏向部分客户端的问题，因此 FedLF 需要通过专门设计的优化问题获取各层更新方向片段 $\mathbf{d}_l^t$ ，即第 t 轮 l 层的更新方向。具体为求解最小化问题：

$$\begin{cases} (\mathbf{d}_l^t, \alpha_l^t) = \arg \min_{\substack{\mathbf{d}_l^t \in \mathbf{R}^{n_l}, \alpha_l^t \in \mathbf{R}}} \alpha_l^t + \frac{1}{2} \|\mathbf{d}_l^t\|^2 \\ \text{s.t. } \mathbf{g}_{i,l}^t \cdot \mathbf{d}_l^t \leq \alpha_l^t, \forall i = 1, 2, \dots, m \\ \mathbf{g}_{p,l}^t \cdot \mathbf{d}_l^t \leq \alpha_l^t \end{cases} \quad (12)$$

其中， $\mathbf{g}_{p,l}^t$ 为公平驱动梯度 $\mathbf{g}_p^t$ 在层 l 的片段。考虑部分层参数维度过大，直接求解式(12)的效率极低，因此本文基于卡罗需-库思-塔克 (karush-kuhn-tucker, KKT) 条件^[31]推导得到替代表达式：

$$\mathbf{d}_l^t = -\left(\sum_{i=1}^m \lambda_i \mathbf{g}_{i,l}^t + \mu \mathbf{g}_{p,l}^t \right) \quad (13)$$

其中， λ_i 和 μ 需要满足非负性及求和为 1 的约束，这些参数可以通过求解下面的 $(m+1)$ 维对偶二次优化问题高效获取：

$$\begin{cases} \max_{\lambda_i, \mu} -\frac{1}{2} \left\| \sum_{i=1}^m \lambda_i \mathbf{g}_{i,l}^t + \mu \mathbf{g}_{p,l}^t \right\|^2 \\ \text{s.t. } \sum_{i=1}^m \lambda_i + \mu = 1 \\ \lambda_i, \mu \geq 0, \forall i = 1, 2, \dots, m \end{cases} \quad (14)$$

受文献[32]启发得出，层级方向片段 $\mathbf{d}_l^t$ 满足

以下特性：若存在非负向量 $\xi \in \mathbf{R}^{m+1}$ （且 ξ 各元素之和为1），使得 $\sum_{i=1}^m \mathbf{g}_{i,l}^t \xi_i + \mathbf{g}_{p,l}^t \xi_{m+1} = \mathbf{0}$ ，则 $\mathbf{d}_l^t = \mathbf{0}$ ；若所有层的 $\mathbf{d}_l^t$ 均为 $\mathbf{0}$ ，则全局更新方向 $\mathbf{d}^t = \mathbf{0}$ ，此时模型 ω^t 达到帕累托平稳状态；但如果仅部分层的 $\mathbf{d}_l^t = \mathbf{0}$ ，这些层的参数会陷入停滞、停止更新，这种情况通常发生在梯度矩阵 $(\mathbf{g}_{1,l}^t, \dots, \mathbf{g}_{m,l}^t, \mathbf{g}_{p,l}^t)$ 的秩小于 m 时。当层参数维度 $n_l \geq m$ 时，该梯度矩阵的秩几乎必然等于 m ；但部分层的参数维度 $n_l < m$ 时，梯度矩阵秩可能小于 m ，进而导致 $\mathbf{d}_l^t = \mathbf{0}$ 。为解决这一问题，当某层的 $n_l < m$ 时，本文将该层与相邻层合并，并重新计算合并后层的 $\mathbf{d}_l^t$ ；若结果仍为 $\mathbf{0}$ ，则重复层合并操作，直到得到非零方向或所有层合并完成。最后合并所有层得到模型更新方向 $\mathbf{d}^t$ 。随后对 $\mathbf{d}^t$ 归一化处理 $\mathbf{d}^t = \frac{\mathbf{d}^t}{\|\mathbf{d}^t\|} \cdot \|\bar{\mathbf{d}}\|$ ，其中 $\bar{\mathbf{d}} = -\frac{1}{m} \sum_{i=1}^m \mathbf{g}_i^t$ 。这样得到的更新方向 $\mathbf{d}^t$ 可以在降低公平性驱动目标的同时应对模型级冲突和层级冲突。

Non-IID 数据分布会干扰联邦学习的公平性，导致局部模型在训练过程中逐渐偏离全局最优方向，不仅降低了全局模型精度，还进一步损害了系统公平性。因此本文采用 ADMM 优化损失函数以缓解这一影响，如下：

$$\mathcal{L}_i(\omega_i^t, \mathbf{y}_i^t, \omega^t) = f_i(\omega_i^t) + (\mathbf{y}_i^t)^T (\omega_i^t - \omega^t) + \frac{\rho}{2} \|\omega_i^t - \omega^t\|^2 \quad (15)$$

其中， $\mathbf{y}_i^t \in \mathbf{R}^d$ 是客户端 i 持有的本地对偶变量， t 为当前轮次， $\rho > 0$ 是二次项系数。

客户端 i 通过最小化该局部损失函数来更新模型：

$$\omega_i^{t+1} \approx \arg \min_{\omega_i^t} \mathcal{L}_i(\omega_i^t, \mathbf{y}_i^t, \omega^t) \quad (16)$$

接着，更新对偶变量 $\mathbf{y}_i^t$ ：

$$\mathbf{y}_i^{t+1} = \mathbf{y}_i^t + \rho(\omega_i^{t+1} - \omega^t) \quad (17)$$

最后，服务器完成全局模型的更新：

$$\omega^{t+1} = \omega^t + \frac{\eta}{|C^t|} \left((\omega_i^{t+1} - \omega^t) + \frac{1}{\rho} (\mathbf{y}_i^{t+1} - \mathbf{y}_i^t) \right) \quad (18)$$

其中， C^t 是第 t 轮被选中参与本地训练的客户端集合， η 是全局学习率。本文依据 FedADMM^[33] 的经验，将 ρ 的值设为 1，该取值可平衡 ADMM

优化过程中的约束满足度与目标函数优化速度，既避免 ρ 过小导致一致性约束失效、局部模型偏离全局最优方向，也防止因 ρ 过大让惩罚项占比过高、压制局部模型的个性化更新，是分布式场景下 ADMM 算法收敛性与优化效率的理论平衡解。

自适应 LDP 技术的核心逻辑实现如下：所有操作均在客户端本地执行，基于模型分层权重特征实现个性化扰动。当客户端 i 获得第 t 轮本地模型 ω_i^t ，它首先计算每一层的权重范围，客户端 i 计算第 l 层权重最大值 $\omega_{\max}^l$ 和最小值 $\omega_{\min}^l$ ，计算该层的中心 c_l 与半径 r_l ：

$$\begin{cases} c_l = \frac{\omega_{\max}^l + \omega_{\min}^l}{2} \\ r_l = \omega_{\max}^l - c_l \end{cases} \quad (19)$$

为确保该层所有权重落在 $[c_l - r_l, c_l + r_l]$ 区间内，本文对该层内的每个权重 ω 计算其相对于层中心 c_l 的偏移量，作为扰动的核心对象：

$$\mu = \omega - c_l \quad (20)$$

其中，偏移量 μ 可正可负，且满足 $|\mu| < r_l$ ，确保扰动始终围绕层中心进行，不偏离权重分布范围。为了进一步减少扰动范围，本文将扰动函数 $\mathcal{M}$ 设计为分段函数，根据不同条件以不同方式扰动不同的权重。为此，首先从伯努利分布中随机采样一个伯努利变量 b ，其满足指定的概率，伯努利变量 b 取 1 的概率如下：

$$\Pr[b = 1] = \frac{e^\epsilon - 1}{2e^\epsilon} \quad (21)$$

基于采样得到的伯努利变量 b ，对偏移量 μ 进行自适应缩放，生成扰动后权重 ω^* ：

$$\mathcal{M}(\omega) = \omega^* = \begin{cases} c_l + \mu \cdot \frac{e^\epsilon + 1}{e^\epsilon - 1}, & b = 1 \\ c_l + \mu \cdot \frac{e^\epsilon - 1}{e^\epsilon + 1}, & b = 0 \end{cases} \quad (22)$$

该函数通过调整偏移量的缩放比例实现隐私保护，既不改变层中心位置，又能根据权重个体特征差异化扰动。文献[30]用半径 r_l 代替了偏移量 μ ，其缩放因子也略有不同，如下：

$$\mathcal{M}(\omega) = \omega^* = \begin{cases} c_l + r_l \cdot \frac{e^\epsilon + 1}{e^\epsilon - 1}, & b = 1 \\ c_l - r_l \cdot \frac{e^\epsilon + 1}{e^\epsilon - 1}, & b = 0 \end{cases} \quad (23)$$

在该方案中, 半径 r_l 和缩放因子是固定的, 而本文则根据每个权重相对于中心 c_l 的偏移量进行扰动, 使得同一层中的所有权重被扰动为不同的值, 从而实现了真正的自适应性。由于 μ 可正可负, 为简单起见, 假设偏移量 μ 为正; 当 μ 为负时, 结论同样成立。因此本文方案中添加的噪声量为:

$$\text{ADD}_{\text{ours}} = \begin{cases} \mu \cdot \frac{e^\epsilon + 1}{e^\epsilon - 1} - \mu, & b = 1 \\ \mu - \mu \cdot \frac{e^\epsilon - 1}{e^\epsilon + 1}, & b = 0 \end{cases} \quad (24)$$

而在文献[30]中, 添加的噪声量为:

$$\text{ADD} = \begin{cases} r_l \cdot \frac{e^\epsilon + 1}{e^\epsilon - 1} - \mu, & b = 1 \\ \mu + r_l \cdot \frac{e^\epsilon + 1}{e^\epsilon - 1}, & b = 0 \end{cases} \quad (25)$$

由于 $|\mu| < r_l$, 通过比较式(24)和式(25)可以得出, 本文的方案引入了更少的噪声。

对于任意相邻权重 ω 和 ω' , 其偏移量分别为 μ 和 μ' , 其概率比值需满足:

$$\frac{\Pr[\mathcal{M}(\omega) = Y]}{\Pr[\mathcal{M}(\omega') = Y]} \leq \frac{e^\epsilon + 1}{e^\epsilon - 1} \leq e^\epsilon \quad (26)$$

由式(26)解得 $\epsilon \geq \ln(1 + \sqrt{2})$ 。故当隐私预算满足该条件时, 机制满足 ϵ -LDP。本文实验中将 ϵ 取为 2, 已满足上述阈值。本文的实验结果验证了当 ϵ 取值为 2 时, 在实现隐私保护的同时仅略微降低准确率。

本文算法的核心在于通过 FedLF 的分层公平优化与 ADMM 的一致性约束的深度融合, 实现公平性与模型性能的联合提升。具体而言, FedLF 从梯度层面缓解了模型级与层级的梯度冲突, 为公平性提供基础优化方向; 而 ADMM 则从优化结构层面引入对偶变量与二次惩罚项, 将局部与全局模型的一致性显式编码到客户端子问题中, 从而在 Non-IID 数据下抑制局部训练偏离全局最优, 避免公平性在局部迭代中被削弱。二者形成“方向引导+结构约束”的互补机制: FedLF 确保更新方向在不同客户端之间均衡, ADMM 则保障局部更新过程始终与全局目标对齐, 从而在复杂异构场景下同时提升公平性与模型性能。在此基础上, LDP 模块以轻量级自适应扰动方式在模型上传前进行隐私保护, 其扰动粒

度与模型层级结构对齐, 使得隐私保护机制可无缝嵌入上述优化流程中, 不干扰公平性约束与一致性约束的收敛行为。三者分别从公平性优化、一致性约束与隐私保护 3 个维度协同作用, 共同构成面向异构数据的公平隐私保护联邦学习整体框架。

4.2 复杂度分析

从计算复杂度角度分析, 本文方案在标准联邦平均算法的基础上引入了 3 项主要开销: 分层优化、ADMM 一致性约束与 LDP 扰动。FedAvg 的客户端计算复杂度为 $O(E \cdot d)$, 其中 E 为本地训练轮次, d 为模型参数量。服务器聚合复杂度为 $O(m \cdot d)$, 其中 m 为参与客户端数。本文方案中, 分层优化的额外计算主要来自每轮对每层求解一个 $m+1$ 维对偶二次优化问题, 其复杂度为 $O(L \cdot m^2)$, 其中 L 为模型层数, m 为本轮参与客户端数量, 该开销远低于神经网络前向与反向传播的计算量 $O(E \cdot d)$ 。ADMM 模块在客户端本地增加了对偶变量更新与二次惩罚项计算, 复杂度为 $O(d)$, 相当于额外一次参数更新的计算量。LDP 扰动仅对上传模型进行逐层缩放处理, 复杂度为 $O(d)$, 且仅在每轮通信前执行一次。因此本文方案主导时间复杂度的部分仍然是 $O(E \cdot d)$, 与 FedAvg 算法的计算复杂度处于同一量级。

在通信开销方面, 本文方案的单轮通信量级仍与 FedAvg 算法保持一致。FedAvg 算法每轮客户端上传本地模型参数, 通信开销为 $O(d)$; 本文方案中, 客户端需要额外上传 ADMM 对应的对偶变量 (与模型参数维度相同, 均为 d), 因此单轮通信开销为 $O(d) + O(d) = O(d)$ 。对偶变量的引入是 ADMM 实现局部与全局一致性约束的必要代价, 通信开销处于同一数量级, 且通信轮次未增加, 整体通信开销仍处于可接受范围内。

4.3 算法流程

本文方案由客户端更新和服务器聚合的主进程两部分组成。算法流程如图 2 所示。

在客户端进程中, 各客户端通过最小化局部目标函数迭代更新本地模型参数。该函数融合私有数据集经验风险、ADMM 对偶变量约束项与强制局部和全局模型之间保持一致性的二次惩罚项, 从而在优化过程中综合考虑多种因素, 在局

部学习目标与全局一致性约束之间实现合理权衡。为兼顾数据隐私保护与模型整体效用，客户端引入LDP隐私机制对更新后的参数添加可控噪声扰动，以降低敏感信息泄露风险。最终将加噪模型参数与局部损失值回传至服务器，完成单轮本地迭代更新。客户端更新如下。

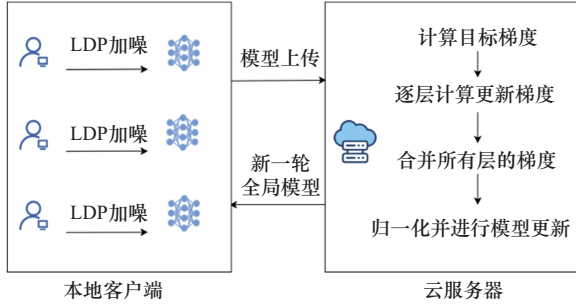


图2 算法流程

算法1 客户端更新

输入 本地学习率 η ，本地训练轮次 E_i ，数据批次集合 B ，局部模型参数 ω_i^t ，本地对偶变量 y_i^t

输出 $\omega_i^{(t+1^*)}, y_i^{(t+1)}, \text{loss } F_i(\omega_i^t)$

- (1) **for** 迭代轮次 $e=0, 1, \dots, E_i-1$ **do**
- (2) **for** 每个批次 $b \in B$ **do**
- (3) 根据式(16)更新局部模型
- (4) **end for**
- (5) **end for**
- (6) **for** $l \in L$ **in** ω_i^{t+1} **do**
- (7) 根据式(19)计算所有层中心 c_l
- (8) **for** $\omega \in \omega_{i,l}^{t+1}$ **do**
- (9) 根据式(20)计算偏移量 μ
- (10) 根据式(21)采样伯努利变量 b
- (11) 根据式(22)对 ω 自适应添加噪声
- (12) **end for**
- (13) **end for**
- (14) 最终得到加噪模型 $\omega_i^{t+1^*}$
- (15) 更新对偶变量 $y_i^{t+1} = y_i^t + \rho(\omega_i^{t+1} - \omega_i^t)$
- (16) **return** $\omega_i^{t+1^*}, y_i^{t+1}, \text{loss } F_i(\omega_i^t)$

在主进程中，服务器首先完成全局模型的初始化，随后进入迭代训练阶段。每一轮通信时，服务器随机采样本轮参与训练的客户端子集，将当前全局模型下发至选中客户端。客户端在本地利用ADMM约束进行优化，完成局部模型更新

后，将模型参数与损失值返回服务器。服务器收集所有本地更新后，借助公平驱动目标计算公平驱动梯度，并逐层计算梯度，以同时缓解模型级与层级梯度冲突；若某层出现零更新方向，则将其与相邻层合并后重新计算。随后，服务器将所有层方向拼接为整体更新方向，经归一化缩放后对全局模型进行更新，如此循环直至模型收敛。整个过程通过分层聚合与ADMM约束，强化了局部与全局模型的一致性，在提升公平性的同时兼顾了模型性能。

算法2 主进程

输入 初始化模型 ω^0 ，学习率 η ，全局轮次 T

输出 全局模型 ω^{t+1}

- (1) **for** 全局迭代轮次 $t = 0, 1, \dots, T-1$ **do**
- (2) 服务器筛选本轮客户端集合 C^t
- (3) **for** $C_i \in C^t$ **do**
- (4) $\omega_i^{t+1^*}, y_i^{t+1}, \text{loss } F_i(\omega_i^t) \leftarrow \text{ClientUpdate}(\omega_i^t)$
- (5) $g_i^t \leftarrow (\omega_i^{t+1^*} - \omega_i^t) + 1/\rho(y_i^{t+1} - y_i^t)$
- (6) **end for**
- (7) 服务器获得本轮所有参与客户端的梯度 g_i^t
- (8) 根据式(11)计算公平性驱动目标梯度
- (9) **for** $l \in L$ **do**
- (10) 根据式(13)计算 d_l^t
- (11) **end for**
- (12) **while** $\exists d_l^t = 0$ **do**
- (13) 合并相邻层重新计算 d_l^t
- (14) **end while**
- (15) 合并所有层得到 d^t
- (16) **if** $\|d^t\| = 0$
- (17) 停止本轮模型训练
- (18) 对 d^t 进行归一化处理
- (19) 更新模型 $\omega^{t+1} \leftarrow \omega^t + \eta d^t$
- (20) **end for**
- (21) **return** ω^{t+1}

5 实验与结果分析

5.1 实验设置

本文使用的公平性指标为问题定义中的公平性指标，具体指的是客户端准确率构成的一维向

量与全 1 向量之间的夹角。本文实验选用 Fashion-MNIST^[34]、CIFAR-10 及 CIFAR-100^[35] 这 3 类公开数据集以覆盖不同复杂度的图像分类任务，其中 Fashion-MNIST 为 28 像素×28 像素的单通道灰度图，包含 10 类衣物图像；CIFAR-10 为 32 像素×32 像素的红绿蓝（RGB）三通道彩色图，涵盖 10 类自然场景图像；CIFAR-100 同样为 32 像素×32 像素的 RGB 三通道彩色图，类别数扩展至 100 类细分类图像，3 类数据集难度依次递增，数据分布更具挑战性，可充分验证算法在高度异构环境中的适配能力和稳定性。对于 Fashion-MNIST 和 CIFAR-10 使用的卷积神经网络（CNN）架构包含两个 5×5 的卷积层、两个 2×2 的最大池化层及两层全连接层。对于 CIFAR-100，用相对复杂的 CNN 模型进行训练，该模型包含 3 个 3×3 的卷积层、3 个 2×2 的最大池化层、批量归一化层、ReLU 激活层、Dropout 层及两层全连接层。

在实验数据预处理阶段，本文采用两种不同的数据集划分模式：第一种为每个客户端含两种类别的数据，此场景为经典 Non-IID 设置划分方法，用于测试算法在常规异质环境下的表现；第二种为每个客户端仅含一种类别的数据，是一种严苛的数据孤岛场景，突出数据分布的极端异构特性，数据异构程度更高，更能考验算法的鲁棒性与性能上限。本文的算法分为 FedALF 及其隐私保护版本 LDP-FedALF，后者在模型上传前引入了自适应 LDP 技术。本文选用的基线方案为

FedLF 算法^[16]、FedAvg 算法^[1]、qFedAvg 算法^[23]，基线方案严格参照原作者的开源代码进行重构，超参数设置与原文最佳状态保持一致，确保对比实验的公平性。本文在硬件环境为单卡 RTX 3090 的云服务器上进行仿真实验。Python 版本为 Python 3.8，PyTorch 版本为 PyTorch 1.13.1。实验的参数设置如下：全局迭代轮数为 500 轮，客户端数量为 100 个，每轮参与客户端的比例为 0.1，批处理大小为 50，全局学习率 η 为 0.5，惩罚系数 ρ 为 1，隐私预算 ϵ 为 2。

5.2 实验结果分析

本文测试了不同算法在 Fashion-MNIST、CIFAR-10 及 CIFAR-100 数据集下的平均测试准确率和公平性指标，所有实验均独立重复运行 10 次，记录均值与标准差，平均测试准确率（公平性指标）见表 3。FedALF 算法和 LDP-FedALF 算法在平均测试准确率和公平性方面较对比方法均有提升。相比 FedALF 算法，LDP-FedALF 算法在实现隐私保护的基础上仅略微降低了平均测试准确率和公平性。Pat1 为每个客户端含两种类别的划分策略；Pat2 为每个客户端仅含一种类别数据的划分策略；公平性指标为客户端平均测试准确率构成的向量与全一向量的夹角，值越小则代表公平性越好。

图 3~图 8 分别展示了平均测试准确率和公平性指标随全局迭代轮数变化的曲线。实验为 Pat1 数据分割下不同数据集实验结果。

表 3 平均测试精度（公平性指标）

数据集	划分	平均测试准确率				
		FedAvg	FedLF	FedALF	qFedAvg	LDP-FedALF
Fashion-MNIST	Pat1	84.28±0.42 (0.171±0.011)	86.38±0.31 (0.122±0.007)	87.12±0.28 (0.109±0.006)	82.72±0.53 (0.163±0.012)	86.46±0.35 (0.116±0.008)
	Pat2	82.51±0.48 (0.204±0.013)	85.21±0.29 (0.158±0.009)	85.58±0.33 (0.130±0.007)	80.53±0.56 (0.196±0.014)	85.27±0.38 (0.137±0.009)
CIFAR-10	Pat1	53.59±0.67 (0.196±0.012)	57.83±0.52 (0.146±0.009)	60.08±0.45 (0.124±0.008)	52.24±0.71 (0.167±0.011)	58.35±0.54 (0.132±0.009)
	Pat2	49.73±0.73 (0.286±0.015)	55.36±0.61 (0.174±0.010)	57.03±0.58 (0.158±0.009)	48.49±0.82 (0.204±0.013)	55.47±0.66 (0.165±0.010)
CIFAR-100	Pat1	29.54±0.88 (0.318±0.017)	32.19±0.76 (0.253±0.013)	35.04±0.69 (0.226±0.012)	27.50±0.97 (0.293±0.016)	33.18±0.74 (0.239±0.013)
	Pat2	26.56±0.94 (0.395±0.019)	30.13±0.82 (0.264±0.014)	33.08±0.73 (0.247±0.013)	25.19±1.03 (0.352±0.018)	31.39±0.79 (0.256±0.014)

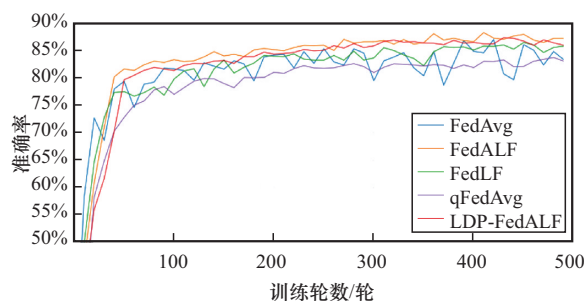


图3 Fashion-MNIST 平均测试准确率对比

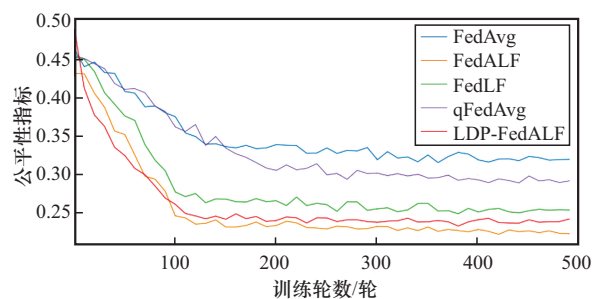


图8 CIFAR-100 公平性指标对比

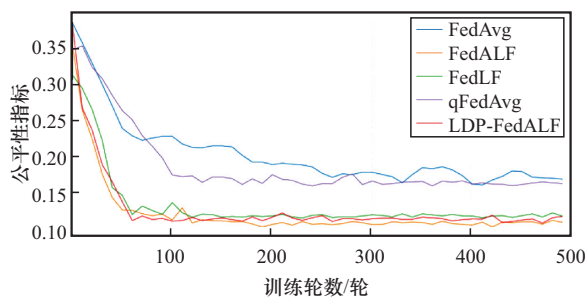


图4 Fashion-MNIST 公平性指标对比

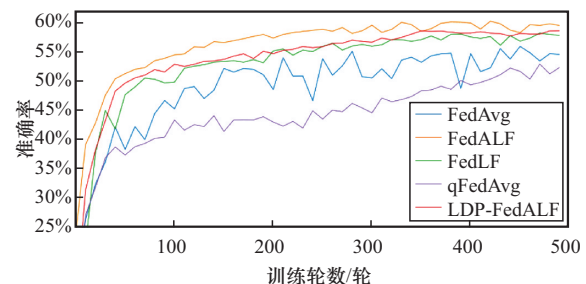


图5 CIFAR-10 平均测试准确率对比

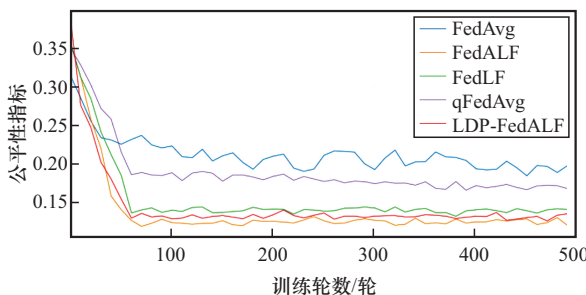


图6 CIFAR-10 公平性指标对比

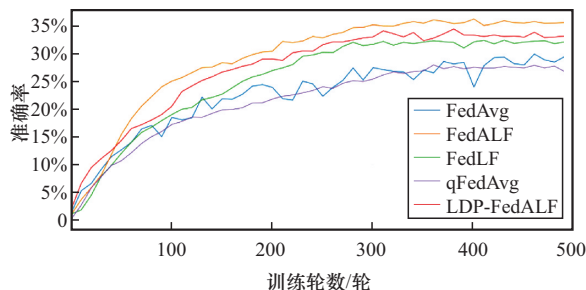


图7 CIFAR-100 平均测试准确率对比

从图3~图8可以清晰地看出，FedALF 算法和 LDP-FedALF 算法的收敛速度均快于 FedAvg 算法、FedLF 算法、qFedAvg 算法等对比联邦学习算法，在数据集的训练初期，这两种算法平均测试准确率快速攀升并率先进入稳定阶段，而其他基线算法仍需更多迭代轮数才能达到相近的收敛状态；且收敛后的准确率始终保持领先优势，尤其是在数据复杂度更高的 CIFAR-100 数据集上，这种精度优势更为明显，充分体现了其在异质数据场景下更强的模型泛化能力。在公平性方面，本文算法的公平性指标下降速度更快，从训练初期便快速低于其他对比算法，且随着迭代轮数的增加持续稳步下降，最终稳定在更低的数值区间。相较于 FedLF 算法，LDP-FedALF 算法在部分数据集的平均测试准确率最高提升 4.18%，公平性最高提升 13.29%。相比于 FedALF 算法，LDP-FedALF 算法在实现隐私保护的基础上略微降低了准确率和公平性。

CIFAR-10 数据集上模型在 10 个客户端下的测试准确率热力图如图9所示。由图9可以看到，之前的方法在某些客户端上性能显著下降，而有利于其他客户端。相比之下，本文算法获得了一个更公平的模型，在客户端之间具有更均匀和更高的精度。

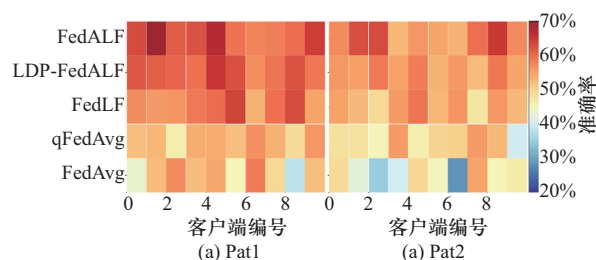


图9 CIFAR-10 数据集上模型在 10 个客户端下的测试准确率热力图

不同隐私保护机制下模型测试准确率对比如图 10 所示。图 10 对比了 FedALF 算法、LDP-FedALF 算法与 DP-FedALF (传统差分隐私) 在相同隐私预算 ($\epsilon=2$) 下的测试准确率变化情况。从图 10 可以看出, 在 Fashion-MNIST 数据集上, LDP-FedALF 的准确率曲线与 FedALF 整体趋势基本相近, 收敛后二者准确率保持在相近水平; 而 DP-FedALF 在训练过程中准确率明显偏低, 收敛后与 FedALF 相差大约 8 个百分点。上述结果表明, 本文提出的自适应 LDP 机制在实现严格隐私保护 (ϵ -LDP) 的同时, 能够将准确率损失控制在较小范围内, 相较于传统差分隐私方案在模型可用性方面具有明显优势。

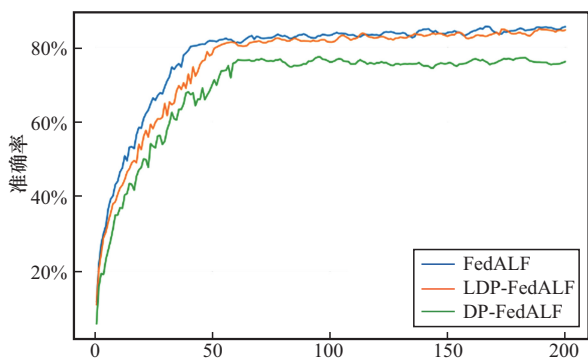


图 10 不同隐私保护机制下模型测试准确率对比

6 结束语

针对 Non-IDD 数据场景下联邦学习面临的模型公平性失衡与参数传输隐私泄露两大核心挑战, 本文聚焦现有公平联邦学习算法在复杂异构环境中局部与全局模型一致性不足、隐私保护缺失的关键问题, 提出了融合交替方向乘子法与自适应 LDP 的 LDP-FedALF 算法。该算法通过静态公平驱动目标与分层优化逻辑, 有效缓解了模型级与层级梯度冲突; 借助 ADMM 优化框架引入对偶变量项与二次惩罚项, 构建局部和全局模型一致性约束, 从根本上抑制了局部训练偏离全局最优方向的问题, 实现了公平性与模型性能的同步提升; 同时, 通过客户端本地执行的自适应 LDP 技术, 在满足隐私安全要求的前提下最大限度地降低了噪声对模型精度的影响。实验结果充分验证了算法的优越性, 在 Fashion-MNIST、CIFAR-10 及 CIFAR-100 这 3 类数据集的不同异构

场景下, 相较于现有公平性联邦学习算法, 本文的算法准确率和公平性都有提升。相比于 FedALF, LDP-FedALF 在实现隐私保护的同时仅带来轻微性能损耗。

未来研究可进一步探索 ADMM 超参数的自适应调整策略和自适应调整公平性驱动目标, 以适配更复杂的系统异构环境。

参考文献:

- [1] McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Artificial Intelligence and Statistics. Fort Lauderdale: PMLR, 2017: 1273-1282.
- [2] 周传鑫, 孙奕, 汪德刚, 等. 联邦学习研究综述[J]. 网络与信息安全学报, 2021, 7(5): 77-92.
- [3] Zhou C X, Sun Y, Wang D G, et al. Survey of federated learning research[J]. Chinese Journal of Network and Information Security, 2021, 7(5): 77-92.
- [4] Hu B T, Lu Z D, Li H, et al. Convolutional neural network architectures for matching natural language sentences[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2. New York: ACM, 2014: 2042-2050.
- [5] Wan F, Wei P X, Han Z J, et al. Min-entropy latent model for weakly supervised object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(10): 2395-2409.
- [6] Lu J W, Liong V E, Zhou X Z, et al. Learning compact binary face descriptor for face recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(10): 2041-2056.
- [7] Li T, Sahu A K, Zaheer M, et al. Federated optimization in heterogeneous networks[C]//Proceedings of Machine Learning and Systems 2020. Austin: mlsys.org, 2020: 429-450.
- [8] Shi Y X, Yu H, Leung C. Towards fairness-aware federated learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(9): 11922-11938.
- [9] Li T, Hu S, Beirami A, et al. Ditto: Fair and robust federated learning through personalization[C]//International conference on machine learning. PMLR, 2021: 6357-6368.
- [10] Mohri M, Sivek G, Suresh A T. Agnostic federated learning[C]//International Conference on Machine Learning, ICML 2019. Long Beach: PMLR, 2019: 4615-4625.
- [11] Wang Z, Fan X L, Qi J Z, et al. Federated learning with fair averaging[PP]. arXiv preprint, 2021, arXiv: 2104.14937.
- [12] Hu Z O, Shaloudegi K, Zhang G J, et al. Federated learning meets multi-objective optimization[J]. IEEE Transactions on Network Science and Engineering, 2022, 9(4): 2039-2051.
- [13] Pan Z B, Wang S Y, Li C, et al. FedMDFG: federated learning with multi-gradient descent and fair guidance[J]. Proceedings of the

- AAAI Conference on Artificial Intelligence, 2023, 37(8): 9364-9371.
- [13] Yu F, Wang D Q, Shelhamer E, et al. Deep layer aggregation[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 2403-2412.
- [14] Ma X S, Zhang J, Guo S, et al. Layer-wised model aggregation for personalized federated learning[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 10082-10091.
- [15] Lee S, Zhang T, Avestimehr S. Layer-wise adaptive model aggregation for scalable federated learning[C]//Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence. New York: ACM, 2023: 8491-8499.
- [16] Pan Z B, Li C, Yu F C, et al. FedLF: layer-wise fair federated learning[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(13): 14527-14535.
- [17] 迟欢欢, 熊平, 刘恒竹, 等. 激励驱动的联邦学习研究综述: 隐私与安全[J]. 网络与信息安全学报, 2025, 11(3): 19-44.
- Chi H H, Xiong P, Liu H Z, et al. Survey on incentive-driven federated learning: privacy and security[J]. Chinese Journal of Network and Information Security, 2025, 11(3): 19-44.
- [18] Zhu L, Liu Z, Han S. Deep leakage from gradients[C]//Advances in Neural Information Processing Systems: Vol. 32. Vancouver: Curran Associates, 2019: 14774-14784.
- [19] 赵瑞, 段小文, 刘新, 等. 非独立同分布数据下层次化 Sinkhorn 距离聚类联邦学习算法[J]. 计算机研究与发展, 2025, 62(12): 3002-3012.
- Zhao R, Duan X W, Liu X, et al. Hierarchical sinkhorn distance-based clustering federated learning algorithm for Non-IID data[J]. Journal of Computer Research and Development, 2025, 62(12): 3002-3012.
- [20] 李宁, 徐家磊, 李菲, 等. 自适应聚合与元学习优化的联邦学习方法研究[J]. 计算机技术与发展, 2026, 36(3): 178-185.
- Li N, Xu J L, Li F, et al. Federated learning with adaptive aggregation and meta-learning optimization[J]. Computer Technology and Development, 2026, 36(3): 178-185.
- [21] Mao W J, Yu B, Zhang C, et al. FedKT: federated learning with knowledge transfer for Non-IID data[J]. Pattern Recognition, 2025, 159: 111143.
- [22] Sun B L, Song X, Tu Y C, et al. FedAgent: federated learning on non-IID data via reinforcement learning and knowledge distillation[J]. Expert Systems with Applications, 2025, 285: 127973.
- [23] Li T, Sanjabi M, Beirami A, et al. Fair resource allocation in federated learning[PP]. arXiv preprint, 2020, arXiv: 1905.10497.
- [24] Ling X P, Fu J, Wang K C, et al. FedFDP: fairness-aware federated learning with differential privacy[PP]. arXiv preprint, 2026, arXiv: 2402.16028.
- [25] Sabah F, Chen Y W, Yang Z, et al. FairDPFL-SCS: fair dynamic personalized federated learning with strategic client selection for improved accuracy and fairness[J]. Information Fusion, 2025, 115: 102756.
- [26] Wei E M, Ozdaglar A. Distributed alternating direction method of multipliers[C]//Proceedings of the 2012 IEEE 51st IEEE Conference on Decision and Control (CDC). Piscataway: IEEE Press, 2013: 5445-5450.
- [27] Hong M Y, Luo Z Q, Razaviyayn M. Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems[C]//Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2015: 3836-3840.
- [28] Abadi M, Chu A, Goodfellow I, et al. Deep learning with differential privacy[C]//Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2016: 308-318.
- [29] Miao Y B, Xie R P, Li X H, et al. Compressed federated learning based on adaptive local differential privacy[C]//Proceedings of the 38th Annual Computer Security Applications Conference. New York: ACM, 2022: 159-170.
- [30] Sun L C, Qian J W, Chen X. LDP-FL: practical private aggregation in federated learning with local differential privacy[PP]. arXiv preprint, 2021, arXiv: 2007.15789.
- [31] Fliege J, Svaiter B F. Steepest descent methods for multicriteria optimization[J]. Mathematical Methods of Operations Research, 2000, 51(3): 479-494.
- [32] Gebken B, Peitz S, Dellnitz M. On the hierarchical structure of Pareto critical sets[J]. Journal of Global Optimization, 2019, 73(4): 891-913.
- [33] Gong Y H, Li Y C, Freris N M. FedADMM: a robust federated deep learning framework with adaptivity to system heterogeneity[C]//Proceedings of the 2022 IEEE 38th International Conference on Data Engineering (ICDE). Piscataway: IEEE Press, 2022: 2575-2587.
- [34] Xiao H, Rasul K, Vollgraf R. Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms[PP]. arXiv preprint, 2017, arXiv: 1708.07747.
- [35] Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images[R]. Toronto: University of Toronto, 2009.

[作者简介]



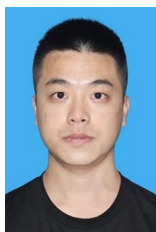
吴炬 (2002-), 男, 西安电子科技大学网络与信息安全学院硕士生, 主要研究方向为联邦学习、隐私保护。



苗银宾 (1988-), 男, 西安电子科技大学网络与信息安全学院教授, 主要研究方向为隐私计算、人工智能安全。



刘淑英 (1982-), 女, 咸阳师范学院计算机学院教授, 主要研究方向为机器学习、网络数据安全。



张伟斌 (2000-), 男, 西安电子科技大学网络与信息安全学院博士生, 主要研究方向为可信人工智能、无线感知安全。



马建峰 (1963-), 男, 西安电子科技大学网络与信息安全学院教授, 主要研究方向为应用密码学、无线网络安全、数据安全、移动智能系统安全。