基于本地化差分隐私的联邦学习方法研究

康海燕，冀源蕊
（北京信息科技大学信息管理学院，北京 100192）

摘 要：联邦学习作为一种协作式机器学习方法，允许用户通过共享模型而不是原始数据进行多方模型训练，在实现隐私保护的同时充分利用用户数据，然而攻击者仍有可能通过窃听联邦学习参与方共享模型来窃取用户信息。为了解决联邦学习训练过程中存在的推理攻击问题，提出一种基于本地化差分隐私的联邦学习（LDP-FL）方法。首先，设计一种本地化差分隐私机制，作用在联邦学习参数的传递过程中，保证联邦模型训练过程免受推理攻击的影响。其次，提出并设计一种适用于联邦学习的性能损失约束机制，通过优化损失函数的约束范围来降低本地化差分隐私联邦模型的性能损失。最后，在 MNIST 和 Fashion MNIST 数据集上通过对比实验验证了所提方法的有效性。

关键词：差分隐私；联邦学习；深度学习
中图分类号：TP309.2
文献标志码：A
DOI: 10.11959/j.issn.1000−436x.2022189

Research on federated learning approach based on local differential privacy

KANG Haiyan, JI Yuanrui
School of Information Management, Beijing Information Science and Technology University, Beijing 100192, China

Abstract: As a type of collaborative machine learning framework, federated learning is capable of preserving private data from participants while training the data into useful models. Nevertheless, from a viewpoint of information theory, it is still vulnerable for a curious server to infer private information from the shared models uploaded by participants. To solve the inference attack problem in federated learning training, a local differential privacy federated learning (LDP-FL) approach was proposed. Firstly, to ensure the federated model training process was protected from inference attacks, a local differential privacy mechanism was designed for transmission of parameters in federated learning. Secondly, a performance loss constraint mechanism for federated learning was proposed and designed to reduce the performance loss of local differential privacy federated model by optimizing the constraint range of the loss function. Finally, the effectiveness of proposed LDP-FL approach was verified by comparative experiments on MNIST and Fashion MNIST datasets.

Keywords: differential privacy, federated learning, deep learning

0 引言

近年来，人工智能技术给人们的生活带来了极大的便利，尤其是机器学习中深度学习这一分支已经广泛应用于图像处理、自然语言处理、语音识别和网络空间安全等领域。为了获得应用效果更好的模型，可以通过增加训练数据量来实现，然而随着训练数据量的增加，隐私泄露的风险也相应提高。研究表明，

收稿日期：2022−05−07；**修回日期**：2022−09−23
基金项目：国家社会科学基金资助项目（No.21BTQ079）；国家自然科学基金资助项目（No.61370139）；教育部人文社科基金资助项目（No.20YJAZH046）；北京未来区块链与隐私计算高精尖创新中心基金资助项目
Foundation Items: The National Social Science Foundation of China (No.21BTQ079), The National Natural Science Foundation of China (No.61370139), The Ministry of Education of Humanities and Social Science Project (No.20YJAZH046), Beijing Advanced Innovation Center for Future Blockchain and Privacy Computing Fund

针对深度学习模型发起的隐私攻击将导致训练数据的隐私泄露，隐私问题限制了深度学习的进一步发展。

联邦学习[1]是解决深度学习隐私问题的突破性技术。联邦学习的逻辑结构与分布式学习相似，即拥有不同训练数据的多个参与方共同执行一个深度学习任务，两者的区别在于联邦学习没有数据收集阶段，而分布式学习需要对数据进行收集，然后将数据分发给多个服务器，再由中央服务器协调进行迭代，从而训练出最终模型。联邦学习通过在各个客户端本地进行学习得到子模型，再交由中心服务器聚合得到最终模型。联邦学习相关的技术和开放性问题在近些年引起了人们的广泛关注[2]，联邦学习与区块链等新兴技术的融合也是目前的研究热点[3-4]。相比于传统的集中式机器学习方法，联邦学习通过在本地进行训练有效降低了数据隐私泄露的风险，然而这并不代表它能完全防御外部隐私攻击。刘艺璇等[5]根据联邦学习的架构将其面临的隐私攻击分为内部攻击和外部攻击，与外部攻击者相比，内部攻击者具备更强大的能力，其不仅可以在训练过程中对梯度或模型参数发起攻击，还能通过替换样本、更改梯度等方式影响模型训练过程。Song 等[6]指出通过对抗攻击，可以从联邦学习参与方所传递的参数中重构出原始的训练数据，从而导致隐私泄露。

针对联邦学习所面临的隐私风险，目前学术界有 2 种解决思路，分别是加密方法和扰动方法。加密方法通过结合密码学工具为联邦训练过程中数据的传输提供隐私保证，常用密码学工具有同态加密和秘密分享。Liu 等[7]设计了一种基于同态加密技术的参数加密方案，抵御联邦学习过程中的投毒攻击。Phong 等[8]利用加法同态加密技术为客户−服务器架构的联邦训练提供保护，然而该算法仅关注本地参数的隐私性，全局梯度对所有终端直接可见。Ou 等[9]设计了一种由第三方掌握私钥、终端利用公钥实现加法同态加密的方案，应用到纵向联邦学习的线性回归模型中实现隐私保护。由于同态加密技术的计算代价昂贵，因此在实践中不适用于大规模数据参与的模型迭代训练。为了在降低计算代价的同时保证中间参数不被泄露，Zhu 等[10]利用秘密分享技术确保至少 t 个用户上传参数后，中心服务器才能进行解密，实现对中间参数的保护。应用秘密分享技术的联邦学习方案虽然不需要大量计算，但增加了通信次数，因此也增加了联邦学习的通信成本。

扰动方法通过差分隐私等技术在模型训练过程中添加噪声扰动，使发布的模型在保持可用性的同时得到保护。差分隐私作为一种轻量级的隐私保护技术[11]，在联邦学习隐私保护领域得到了广泛关注。根据联邦学习中保护对象的不同，可以将扰动方法分为中心化扰动和本地化扰动。中心化扰动主要保护联邦学习中心服务器在获取和下发中间参数时的隐私性。Geyer 等[12]首次提出差分隐私中用户级别联邦学习（CL-FL, client level federated learning）的差分隐私保护方法，通过在服务器端引入高斯噪声来隐藏单个参与方对联邦训练的贡献。为了提高隐私预算利用率，使用矩累计[13]方法获取更紧致的隐私损失边界，然而 Geyer 等在计算隐私损失时直接对梯度进行裁剪的做法浪费了一部分隐私预算。Zhou 等[14]在 CL-FL 的基础上进一步完善了用户级别的隐私保护方法，在提高通信效率的基础上保证了中心参数服务器的隐私性。Wei 等[15]提出了一种分阶段的差分隐私聚合前噪声联邦学习（NbAFL, noise before aggregation federated learning）方法，并证明通过适当调整噪声的方差可以满足不同隐私保护水平下的差分隐私。该方法全面考虑了中心参数传递过程中不同阶段的隐私问题，但需要经过多次迭代才能达到较高的模型准确率。上述中心化扰动方法中的噪声均由中心服务器添加，然而中心参数服务器也可能是半诚实甚至恶意的，因此需要研究本地化扰动方法，本地化扰动方法通常结合本地化差分隐私技术来实现。Truex 等[16]在对联邦学习的参数进行本地化差分隐私扰动时引入 α -CLDP 方法，根据输入样本对的距离分配隐私预算，以较大概率输出与原始值相近的扰动值。由于联邦学习中梯度或模型参数的维度很高，直接扰动会带来很大的通信量，为了提高通信效率，Liu 等[17]提出一种两阶段方法，根据指数机制选择权重最高的 k 个维度的梯度数据，再对所选择的维度数据进行扰动，解决联邦学习中梯度导致隐私泄露问题，并设计 3 种隐私维度选择机制。Zhao 等[18]将梯度数据扰动后的值离散到偶数区间内，通过两位数值即可表示输出值，节约了通信开销，然而这种做法对联邦模型的性能造成了损失。

表 1 对现有研究方案进行了总结，通过表 1 可知，现有研究主要存在如下不足：1) 基于同态加密的联邦学习隐私保护方法计算开销大，基于秘密分享的联邦学习隐私保护方法通信开销太大；2) 中心化扰动方法依赖可信的中心服务器；3) 本地化扰动

表 1 现有研究方案对比

分类	方案	概述	不足
加密	文献[7]	基于同态加密思想设计联邦学习参数保护方法	计算代价高
	文献[8]	由终端自行生成密钥，联邦模型训练时结合加法同态技术对参数进行保护	计算代价高
	文献[9]	引入第三方进行私钥分配后再使用同态加密技术	计算代价高
	文献[10]	结合秘密分享技术对参与方传递的参数进行保护	通信开销大
中心化扰动	文献[12]	设计用户级别的差分隐私联邦学习算法	依赖中心参数服务器
	文献[14]	利用高斯噪声机制对参数进行保护	依赖中心参数服务器
	文献[15]	设计分阶段差分隐私联邦学习模型 NbAFL	依赖中心参数服务器
本地化扰动	文献[16]	利用指数机制思想设计差分隐私参数扰动	性能损失较大
	文献[17]	根据权重对梯度进行选择后再进行扰动	性能损失较大
	文献[18]	选择部分梯度后将结果进行离散后再添加噪声	性能损失较大

方法在模型性能上损失较大，需要从隐私机制设计的角度进行改进。

针对以上不足，本文主要贡献如下。

1) 提出一种基于本地化差分隐私的联邦学习（LDP-FL, local differential privacy federated learning）方法，解决联邦学习训练过程中存在的隐私问题。

2) 设计一种本地化差分隐私机制，作用在联邦学习参数传递过程中，通过设计噪声机制，扰动联邦学习所传递的参数，从而增加联邦模型训练的隐私性。

3) 设计一种性能损失更小的估计机制，通过优化损失函数的约束范围来降低引入本地化差分隐私机制后联邦模型的性能损失。

4) 在 MNIST 和 Fashion MNIST 这 2 个真实的数据集上，分别从全局准确率、性能损失和运行时间 3 个方面进行对比实验，与其他算法相比，本文所提方法效果更优。

1 背景知识

1.1 联邦学习

联邦学习是谷歌提出的一种机器学习方法[1]。在一个典型的联邦学习方法中，通常假设有 N 个参与方和一个中心参数服务器，这些参与方通过协作共同训练出一个可用的深度学习模型。在每次训练迭代时，每个参与方共享的是其本地更新后的模型参数而不是本地的训练数据。记每一个参与方 C_i 拥有对应的数据集 D_i，则 $\left|D\right|=\sum_{i\in N}\left|D_i\right|$，全局模型的目标损失函数记作 $L(D,w)$，联邦学习所面临的优化问题为

$$w^*=\arg\min_w L(D,w)=\arg\min_w\sum_{i=1}^{N}L_i(D_i,w) \quad (1)$$

其中，L_i 表示第 i 个参与方的本地损失函数，一般通过本地经验风险最小化过程（如随机梯度下降等）来求解。联邦学习中的经验风险最小化的过程通常包含如下训练步骤。

1) 初始化：由中心参数服务器对需要训练的深度学习模型进行初始化，并广播给所有参与方。

2) 本地模型训练：接收到初始模型参数的参与方使用本地数据对模型进行训练后，将更新参数传递给中心参数服务器。

3) 全局模型聚合：接收到所有参与方传递参数后，中心参数服务器对获得的模型进行聚合后广播。

1.2 本地化差分隐私

本地化差分隐私技术的核心思想是对用户本地数据添加满足本地化差分隐私的扰动噪声，将扰动后数据传输给第三方数据收集者，再通过一系列操作得到有效的结果。由于传统的 ε – 本地化差分隐私过于严格，目前深度学习隐私保护中常用的是宽松差分隐私，定义如下。

定义 1 (ε,δ) – 本地化差分隐私。给定 N 个用户，每个用户对应一条记录，对于隐私机制 $\mathcal{M}$，其定义域为 $\mathrm{Dom}(\mathcal{M})$，值域为 $\mathrm{Ran}(\mathcal{M})$，若隐私机制 $\mathcal{M}$ 在任意两条记录 t,t' $(t,t'\in\mathrm{Dom}(\mathcal{M}))$ 上得到的输出结果 $(o(o\subseteq\mathrm{Ran}(A)))$ 相同，且满足

$$\Pr(\mathcal{M}(t)=o)\leqslant e^{\varepsilon}\Pr(\mathcal{M}(t')=o)+\delta \quad (2)$$

则称隐私机制 $\mathcal{M}$ 满足 (ε,δ) – 本地化差分隐私。

高斯机制是机器学习隐私保护中常用的一种噪声机制，通过给输出结果 $f(t)$ 添加均值为 0、方

差为 $\sigma^2\boldsymbol{I}$ 的高斯噪声实现 (ε,δ) – 本地化差分隐私，即 $\mathcal{M}(t)=f(t)+\mathcal{M}(0,\sigma^2\boldsymbol{I})$。差分隐私中敏感度的含义是单个数据对查询或分析结果的最大影响值，高斯机制具有 L_2 敏感度，表示根据设定的隐私级别所需设置的扰动值上界，高斯机制中函数 $f(t)$ 的 L_2 敏感度为 $\Delta s=\max\limits_{v,v'\in D}\|f(v)-f(v')\|_2$，为了保证给定的高斯噪声分布 $n\sim\mathcal{N}(0,\sigma^2)$ 满足 (ε,δ) – 本地化差分隐私，所选择的高斯分布标准差需要满足 $\sigma\geqslant\dfrac{c\Delta s}{\varepsilon}$，即在 $\varepsilon\in(0,1)$ 的情况下常数 $c\geqslant\sqrt{2\ln\left(\dfrac{1.25}{\delta}\right)}$。本地化差分隐私具有如下 2 个性质。

1) 后置处理免疫性。对于一个输出结果满足差 (ε,δ) – 本地化差分隐私的机制 $\mathcal{M}$，在这个机制的输出结果上进行任何操作都不会造成额外的隐私损失。

2) 序列组合性。对于 k 个满足 (ε_i,δ_i) – 本地化差分隐私的机制 $\mathcal{M}_1,\cdots,\mathcal{M}_i,\cdots,\mathcal{M}_k$，其序列组合满足 $\left(\sum\limits_{i=1}^{k}\varepsilon_i,\sum\limits_{i=1}^{k}\delta_i\right)$ – 本地化差分隐私。

使用高斯机制向机器学习模型添加噪声时会导致模型产生性能损失，性能损失和本地化差分隐私的关系可以通过定义 2 进行说明。

定义 2　尾约束[11]。对于任意 $\varepsilon>0$，当 $\delta=\min\limits_{\lambda}\exp(\alpha_{\mathcal{M}}(\lambda)-\lambda\varepsilon)$ 时，机制 $\mathcal{M}$ 满足 (ε,δ) – 差分隐私。

2　方法设计

2.1　问题的描述

联邦学习通过将用户数据保留在本地降低了用户训练数据隐私泄露的风险，然而联邦学习过程仍然存在一定的安全问题，对于共同参与模型训练的多个参与方以及中心参数服务器，若它们是诚实且好奇的，即这些参与方在联邦学习过程中会遵守模型的训练协议，但互相对对方的私有数据和模型参数是好奇的，在协作期间会不断推理，希望获取更多对方额外的信息，如训练数据和模型参数。为了抵御这样的推理攻击，需要对联邦模型训练过程提供额外的隐私保护机制，因此本文的目标是设计一种满足本地化差分隐私的联邦学习方法，实现在服务器或参与方诚实且好奇的情况下安全有效地训练联邦模型，即保护参与方的私有数据和模型参数不被攻击者恶意推理的同时保证模型训练的精度。

具体来说，在联邦模型训练过程中，假设完成全局联邦模型的训练需要经过 T 次迭代，在每一次迭代过程 t 中，选择 k 个参与方利用本地数据集对下发的初始模型进行训练，每个参与方将训练好的模型更新结果传输给中心参数服务器，为了防止参与方所训练的模型在传输过程中发生隐私泄露，需要设计一种本地化的隐私机制对传输过程中的模型参数进行隐私保护处理，本文所涉及的相关符号和参数如表 2 所示。

表 2　　**相关符号和参数**

符号	含义
N	联邦学习参与方数量
C_i	联邦学习中第 i 个参与方
D	所有参与方的训练数据集之和
D_i	第 i 个参与方对应的数据集
L_i	第 i 个参与方的本地损失函数
ε	本地化差分隐私定义中的隐私预算
δ	本地化差分隐私定义中相关参数
$\mathcal{M}$	本地化差分隐私扰动机制
σ	本地化差分隐私噪声机制方差
q	联邦学习过程中每次迭代时的采样率
w	联邦学习中的模型参数
T	联邦学习总交流轮次
Δs	本地化差分隐私敏感度
p_i	第 i 个参与方的性能损失函数
P	本地化差分隐私整体性能损失
$\alpha(\lambda)$	λ 时刻的性能损失函数

2.2　本地化差分隐私联邦学习方法的设计

为了解决诚实且好奇的中心参数服务器或参与方的存在导致联邦学习中用户本地数据隐私泄露问题，本文提出了一种 LDP-FL 方法，框架如图 1 所示。该方法由一个中心参数服务器和 N 个联邦学习参与方组成，每个联邦学习参与方拥有一个由中心参数服务器下发的初始模型和本地的训练数据集。

LDP-FL 方法的核心思想是在“数据不动算法动，数据可用不可见”的基础上引入本地化差分隐私机制，为联邦训练过程提供额外的隐私保护。具体来说，首先由中心参数服务器生成初始模型，再广播给所选择的联邦学习参与方，参与方接收到初始模型后利用本地数据集对初始模型进行训练，在每个参与方的本地训练的过程中引入本地化差分隐私机制对模型参数进行扰动，通过传输扰动后的

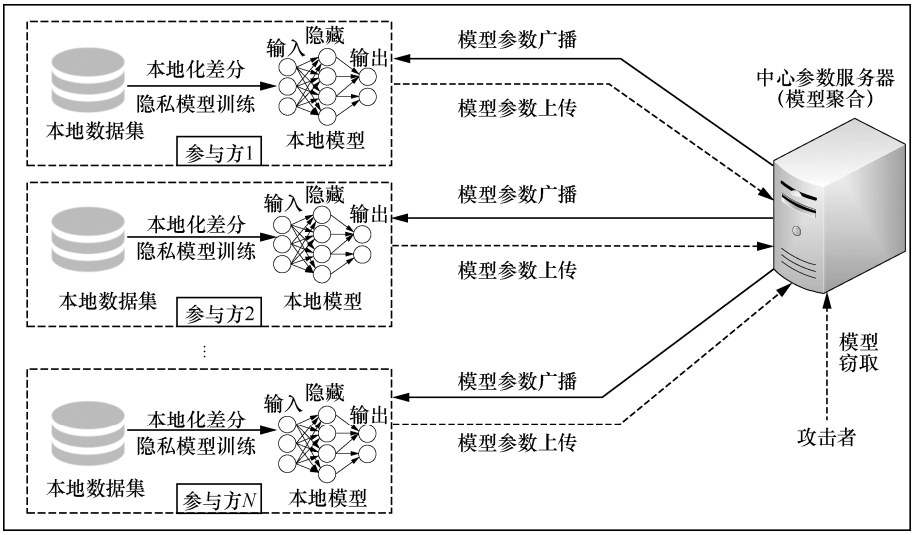


图 1　LDP-FL 方法框架

参数（非原始训练数据）达到隐私保护的目的，中心参数服务器接收到扰动参数后对所有参数进行聚合操作，将聚合后的模型参数再广播给所选择的参与方，不断迭代该过程直到模型收敛。LDP-FL 方法由中心参数服务器处理算法 FL_Server（federated learning server）和参与方本地更新算法 FL_Client（federated learning client）构成。

中心参数服务器处理算法 FL_Server 的具体流程如算法 1 所示。首先，由中心参数服务器对需要训练的模型参数和测试集准确率列表进行初始化。其次，根据设定的迭代次数，在每次迭代时以采样率 q 从 N 个参与方中随机选择 k 个参与方参与训练 $\left(q=\dfrac{k}{N}\right)$，对于所选择的 k 个参与方，将上一轮迭代所获得的全局模型参数 w 传递给算法 2 参与方本地更新算法 FL_Client，k 个参与方以并行化的方式执行该算法，分别获得本次迭代本地模型的参数。最后，当所有参与方完成更新操作后，中心参数服务器对参与方所上传的扰动参数进行聚合处理，即求平均值，获得本次迭代的全局模型参数，使用测试集计算全局模型参数对应的模型准确率，将本轮模型准确率存入测试集准确率列表中，在设定的迭代次数结束后对整体的隐私损失进行估计。

算法 1　FL_Server

输入　联邦学习参与方数量 N，联邦学习采样率 q，联邦学习交流轮次 T

1) 定义列表 test_acc_list，初始化 w_0

2) for　$t \leftarrow 1$　to　T　do

3)　以采样率 q 从 N 个用户中选择 k 个参与方

4)　遍历从 N 中选择的 k 个参与方

5)　$w_{t+1}^{k} \leftarrow \mathrm{FL\_Client}(k, w_t)$

6)　聚合处理：$w_{t+1} \leftarrow \dfrac{1}{n}\sum_{i=1}^{k} w_{t+1}^{i}$

7)　计算本次迭代模型准确率 test_acc

8)　将 test_acc 加入列表 test_acc_list

9) end for

10) 通过 2.3 节性能损失约束机制约束损失函数

11) 返回 test _ acc _ list

算法 2　FL_Client

输入　上一轮训练所得模型参数 w_t，本地模型迭代次数 E，本地数据集大小 m，随机梯度下降中每批次选择的训练集大小 B，随机梯度下降过程学习率 α，本地模型损失函数 $L(w)$，梯度裁剪阈值 C，本地化差分隐私机制隐私参数 ε_i, δ_i

1) for　$e=1$　to　E　do

2)　对于训练集 B 中的每个数据对 b

3)　梯度大小 $g \leftarrow \nabla L(w;b)$

4)　梯度裁剪 $g \leftarrow \dfrac{g}{\max\left(1, \dfrac{\|g\|_2}{C}\right)}$

5)　参数更新 $w^k \leftarrow w - \partial g$

6)　计算敏感度 $\Delta s = \dfrac{2C}{m}$

7)　计算噪声尺度 $\sigma_i = \dfrac{\Delta s\sqrt{2qT\ln\left(\dfrac{1}{\delta_i}\right)}}{\varepsilon_i}$

8)　参数扰动 $\tilde{w}^k \leftarrow w^k + \mathcal{N}(0,\sigma_i^2)$

9) end for

10) 返回扰动参数值 $\tilde{w}^k$

首先，采用随机梯度下降法根据设定的本地迭代轮次 E 对所接收到的初始模型进行训练计算出梯度值，同时引入梯度裁剪技术，目的是限制训练样本对模型参数的影响，通过对梯度的 L2 范数进行裁剪，设定裁剪的阈值为 C，则参与方在每轮本地训练时计算得到的的梯度数据 g_i 将被 $\dfrac{g_i}{\max\left(1,\dfrac{\|g_i\|_2}{C}\right)}$ 替代，梯度裁剪可以保证当 $\|g_i\|_2 \leqslant C$ 时，梯度数据 g_i 被保留；当 $\|g_i\|_2 > C$ 时，梯度数据 g_i 被阈值 C 取代。其次，计算参与方本地训练过程的敏感度，联邦学习中第 i 个参与方的本地训练过程为

$$s^{D_i} = w_i = \arg\min_w L_i(w, D_i) = \frac{1}{|D_i|}\sum_{j=1}^{|D_i|}\arg\min_w L_i(w, D_{i,j}) \tag{3}$$

其中，D_i 表示第 i 个参与方所使用的数据集，$D_{i,j}$ 表示 D_i 中的第 j 个样本。根据本地化差分隐私的定义，考虑 2 个相邻的数据集 D_i 和 D_i'，D_i' 与 D_i 只相差一条数据，则第 i 个客户端本地训练过程 s^{D_i} 的敏感度为

$$\Delta s^{D_i} = \max_{D_i, D_i'} \left\| s_U^{D_i} - s_U^{D_i'} \right\| = \max_{D_i, D_i'} \left\| \frac{1}{|D_i|}\sum_{j=1}^{|D_i|}\arg\min_w L_i(w, D_{i,j}) - \frac{1}{|D_i'|}\sum_{j=1}^{|D_i'|}\arg\min_w L_i(w, D_{i,j}') \right\| = \frac{2C}{|D_i|} \tag{4}$$

根据以上分析，每轮训练的敏感度定义为 $\Delta s = \max\{\Delta s^{D_i}\}$。为了实现最小的全局敏感度，理想情况是每个用户都拥有充足的训练数据，定义本地数据集的大小为 m，则每次迭代时的敏感度 $\Delta s = \dfrac{2C}{m}$。最后，计算所添加高斯噪声的标准差，根据给定的采样率 q 和迭代轮次 T，为了保证联邦训练过程满足 $(\varepsilon_i, \delta_i)$ – 本地化差分隐私，需要给每个参与方训练的模型参数添加一定程度的高斯噪声。给第 i 个参与方本地模型训练所得参数添加高斯噪声的标准差使用式(5)进行计算，在 2.4 节中将具体分析式(5)如何保证 $(\varepsilon_i, \delta_i)$ – 本地化差分隐私。

$$\sigma_i = \frac{\Delta s\sqrt{2qT\ln\left(\dfrac{1}{\delta_i}\right)}}{\varepsilon_i} \tag{5}$$

2.3　性能损失约束机制的设计

通过 2.2 节中的描述可知，引入本地化差分隐私提升联邦训练过程中隐私安全性的同时会给联邦模型的性能造成一定的损失，因此本节设计一种性能损失更小的估计机制，通过这种估计机制降低联邦模型的性能损失。给单个模型添加高斯噪声后的隐私损失需要根据时刻损失函数进行计算，而在联邦学习环境中，需要从 N 个参与方中以采样率 q 选择出 k 个参与方进行联邦模型的训练，记联邦交流的迭代次数为 T，给第 i 个参与方本地模型训练所得的参数添加高斯噪声后每个参与方经过 T 次迭代后隐私损失的计算式为

$$p_i = \exp(\alpha(\lambda)) = \exp\left(\sum_{t=1}^{T}\alpha(\lambda)\right) \tag{6}$$

其中，p_i 表示第 i 个参与方的性能损失。参与联邦训练的 k 个参与方整体的性能损失 P 的计算式为

$$P = \sum_{i=1}^{k} p_i = \sum_{i=1}^{k}\exp(\alpha(\lambda)) = \sum_{i=1}^{k}\exp\left(\sum_{t=1}^{T}\alpha(\lambda)\right) \tag{7}$$

根据文献[13]中的相关定义，将性能函数写作 $\alpha(\lambda) = \log(\max\{E_{v_1,v_0}, E_{v_0,v_1}\})$，其中，$v_0$ 表示高斯分布 $\mathcal{N}(0,\sigma_i^2)$ 的概率密度函数，v_1 表示 $q\mathcal{N}(\Delta s,\sigma_i^2) + (1-q)\mathcal{N}(0,\sigma_i^2)$ 的混合概率密度函数，v_0 和 v_1 分别如式(8)和式(9)所示。

$$v_0(z) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2\sigma_i^2}\right) \tag{8}$$

$$v_1(z) = \frac{q}{\sqrt{2\pi}}\exp\left(-\frac{(z-\Delta s)^2}{2\sigma_i^2}\right) + \frac{(1-q)}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2\sigma_i^2}\right) \tag{9}$$

利用 Renyi 散度对时刻生成函数 $\alpha(\lambda) = \log\left(\max\{E_{v_1,v_0}, E_{v_0,v_1}\}\right)$ 中的 E_{v_1,v_0} 和 E_{v_0,v_1} 做如下定义

$$\begin{cases} E_{v_1,v_0} = \mathbb{E}_{z\sim v_1}\left[\left(\dfrac{v_1(z)}{v_0(z)}\right)^{\lambda}\right] = \mathbb{E}_{z\sim v_0}\left[\left(\dfrac{v_0(z)}{v_1(z)}\right)^{\lambda+1}\right] \\ E_{v_0,v_1} = \mathbb{E}_{z\sim v_0}\left(\dfrac{v_0(z)}{v_1(z)}\right)^{\lambda}\Big] = \mathbb{E}_{z\sim v_0}\left[\left(\dfrac{v_1(z)}{v_0(z)}\right)^{-\lambda}\right] \end{cases} \tag{10}$$

对 E_{v_1,v_0} 和 E_{v_0,v_1} 的大小进行比较，选择更小的一方，并对联邦训练过程的性能损失进行约束。

如式(11)所示构造 $E_{v_1,v_0}-E_{v_0,v_1}$，通过差值的正负来判断两者的大小。

$$E_{v_1,v_0}-E_{v_0,v_1}=\mathbb{E}_{z\sim v_0}\left[\left(\frac{v_0(z)}{v_1(z)}\right)^{\lambda+1}\right]-\mathbb{E}_{z\sim v_0}\left[\left(\frac{v_1(z)}{v_0(z)}\right)^{-\lambda}\right]=$$

$$\int_{-\infty}^{\infty}v_0\left(q\exp\left(\frac{2z\Delta s-\Delta s^2}{2\sigma_i^2}\right)+1-q\right)^{\lambda+1}\mathrm{d}z-$$

$$\int_{-\infty}^{\infty}v_0\left(q\exp\left(\frac{2z\Delta s-\Delta s^2}{2\sigma_i^2}\right)+1-q\right)^{-\lambda}\mathrm{d}z\overset{z=y+\frac{\Delta s}{2}}{=}$$

$$\int_{-\infty}^{\infty}\exp\frac{\left(y+\frac{\Delta s}{2}\right)^2}{2\sigma_i^2}\left(q\exp\left(\frac{y\Delta s}{\sigma_i^2}\right)+1-q\right)^{\lambda+1}\mathrm{d}y-$$

$$\int_{-\infty}^{\infty}\exp\frac{\left(y+\frac{\Delta s}{2}\right)^2}{2\sigma_i^2}\left(q\exp\left(\frac{y\Delta s}{\sigma_i^2}\right)+1-q\right)^{-\lambda}\mathrm{d}y \tag{11}$$

将式(11)中积分的负项进行变换，得到

$$E_{v_1,v_0}-E_{v_0,v_1}=\int_{-\infty}^{0}\exp\frac{\left(y+\frac{\Delta s}{2}\right)^2}{2\sigma_i^2}\left(q\exp\left(\frac{y\Delta s}{\sigma_i^2}\right)+1-q\right)^{\lambda+1}\mathrm{d}y-$$

$$\int_{-\infty}^{0}\exp\frac{\left(y+\frac{\Delta s}{2}\right)^2}{2\sigma_i^2}\left(q\left(\exp\left(\frac{y\Delta s}{\sigma_i^2}\right)-1\right)+1\right)^{-\lambda}\mathrm{d}y\overset{y=-z}{=}$$

$$\int_{0}^{+\infty}\exp\frac{\left(z-\frac{\Delta s}{2}\right)^2}{2\sigma_i^2}\left(q\exp\left(\frac{-z\Delta s}{\sigma_i^2}\right)+1-q\right)^{\lambda+1}\mathrm{d}z-$$

$$\int_{0}^{+\infty}\exp\frac{\left(z-\frac{\Delta s}{2}\right)^2}{2\sigma_i^2}\left(q\left(\exp\left(\frac{-z\Delta s}{\sigma_i^2}\right)-1\right)+1\right)^{-\lambda}\mathrm{d}z \tag{12}$$

对式(12)进行改写得到

$$E_{v_1,v_0}-E_{v_0,v_1}=\int_0^{+\infty}(\Phi_+(z)-\Phi_-(z))\mathrm{d}z \tag{13}$$

将式(13)中 $\Phi_+(z)$ 与 $\Phi_-(z)$ 分别写作

$$\Phi_+(z)=\exp\frac{-\left(z+\frac{\Delta s}{2}\right)^2}{2\sigma_i^2}\left(q\exp\left(\frac{z\Delta s}{\sigma_i^2}\right)+1-q\right)^{\lambda+1}+$$

$$\exp\frac{-\left(z-\frac{\Delta s}{2}\right)^2}{2\sigma_i^2}\left(q\exp\left(\frac{-z\Delta s}{\sigma_i^2}\right)+1-q\right)^{\lambda+1} \tag{14a}$$

$$\Phi_-(z)=\exp\frac{-\left(z+\frac{\Delta s}{2}\right)^2}{2\sigma_i^2}\left(q\exp\left(\frac{z\Delta s}{\sigma_i^2}\right)+1-q\right)^{-\lambda}+$$

$$\exp\frac{-\left(z-\frac{\Delta s}{2}\right)^2}{2\sigma_i^2}\left(q\exp\left(\frac{-z\Delta s}{\sigma_i^2}\right)+1-q\right)^{-\lambda} \tag{14b}$$

记 $\varphi(z)=\dfrac{\Phi_+(z)}{\Phi_-(z)}$，为了确定 $\varphi(z)$ 的单调性，定义 $\theta=\exp\left(\dfrac{z\Delta s}{\sigma_i^2}\right)$，有式(15)成立。

$$\frac{\mathrm{d}\varphi(z)}{\mathrm{d}z}=\frac{\mathrm{d}\varphi(\theta)}{\mathrm{d}\theta}\frac{\mathrm{d}\theta(z)}{\mathrm{d}z}=$$

$$\frac{\mathrm{d}\theta(z)}{\mathrm{d}z}\frac{1-q}{\phi_-^2(z)}(1-q+q\theta)^{\lambda}\left(1-q+\frac{q}{\theta}\right)^{-\lambda-1}\cdot$$

$$\left((\lambda+1)q\left(\theta-\frac{1}{\theta}\right)-(1-q+q\theta)\right)+$$

$$\frac{\mathrm{d}\theta(z)}{\mathrm{d}z}\frac{1-q}{\phi_-^2(z)}\left(1-q+\frac{q}{\theta}\right)^{\lambda}(1-q+q\theta)^{-\lambda-1}\cdot$$

$$\left(\lambda q\left(\theta-\frac{1}{\theta}\right)+(1-q+q\theta)\right) \tag{15}$$

令 $\omega=\dfrac{(\lambda+1)q\left(\theta-\frac{1}{\theta}\right)}{1-q+q\theta}$ 并且考虑 $\omega<1$ 的情况，有式(16)成立。

$$\frac{\omega(1-q+q\theta)}{\lambda+1}=1-q+q\theta-\left(1-q+\frac{q}{\theta}\right) \tag{16}$$

通过等式变换可得到

$$\frac{1-q+\frac{q}{\theta}}{1-q+q\theta}=1-\frac{\omega}{\lambda+1} \tag{17}$$

根据以上相关定义，将 $\dfrac{\mathrm{d}\varphi(z)}{\mathrm{d}z}$ 重写为

$$\frac{\mathrm{d}\varphi(z)}{\mathrm{d}z}=\frac{\mathrm{d}\theta(z)}{\mathrm{d}z}\frac{1-q}{\phi_-^2(z)}\cdot$$

$$(1-q+q\theta)^{\lambda+1}\left(1-q+\frac{q}{\theta}\right)^{-\lambda-1}\cdot$$

$$\left(\omega+1+\left(\frac{\lambda\omega}{\lambda+1}+1\right)\left(1-\frac{\omega}{\lambda+1}\right)^{2\lambda+1}\right) \tag{18}$$

定义 $\psi(\omega)=\left(\omega+1+\left(\dfrac{\lambda\omega}{\lambda+1}+1\right)\left(1-\dfrac{\omega}{\lambda+1}\right)^{2\lambda+1}\right)$，对 $\psi(\omega)$ 分别求一阶导数和二阶导数

$$\begin{aligned}\frac{\mathrm{d}\psi(\omega)}{\mathrm{d}\omega}&=1-\left(1+\frac{2\lambda\omega}{\lambda+1}\right)\left(1-\frac{\omega}{\lambda+1}\right)^{2\lambda+1}\\ \frac{\mathrm{d}^2\psi(\omega)}{\mathrm{d}\omega^2}&=\frac{2\lambda(2\lambda+1)\omega}{(\lambda+1)^2}\left(1-\frac{\omega}{\lambda+1}\right)^{2\lambda-1}\end{aligned} \tag{19}$$

由 $\dfrac{\mathrm{d}\psi(\omega)}{\mathrm{d}\omega}\Big|_{\omega=0}=0$，可知 $\dfrac{d^2\psi(\omega)}{d\omega^2}\geqslant 0$，$\varphi(0)=1$，

由此可得 $\varphi(z)\geqslant 1$，即可得 $E_{v_1,v_0}\geqslant E_{v_0,v_1}$。

通过以上分析可知，式(10)中 E_{v_1,v_0} 与 E_{v_0,v_1} 间存在 $E_{v_1,v_0}\geqslant E_{v_0,v_1}$ 的关系，根据这一关系，对联邦训练过程中 λ 时刻的损失函数进行进一步约束，即 $\alpha(\lambda)=\log E_{v_1,v_0}$，该式说明本文 LDP-FL 方法的损失只需通过 E_{v_1,v_0} 约束即可，从而降低联邦学习过程中的性能损失。

3　隐私安全性与性能分析

3.1　隐私安全性分析

本节对 LDP-FL 方法的隐私安全性进行分析，对于采样率为 q、迭代轮次为 T 的 LDP-FL 方法，有定理 1 成立。

定理 1　为了保证联邦训练中参与方传递模型的过程满足 (ε_i,δ_i) – 本地化差分隐私，所添加的高斯噪声的标准差应满足

$$\sigma_i=\frac{\Delta s\sqrt{2qT\ln\left(\dfrac{1}{\delta_i}\right)}}{\varepsilon_i} \tag{20}$$

证明　利用三角不等式对 E_{v_1,v_0} 进行如下变换

$$E_{v_1,v_0}=\mathbb{E}_{z\sim v_1}\left[\left(v_1\frac{(z)}{v_0(z)^{\lambda}}\right)\right]=$$

$$\int_{-\infty}^{\infty}v_0\left(q\exp\left(\frac{2z\Delta s-\Delta s^2}{2\sigma_i^2}\right)+1-q\right)^{\lambda+1}\mathrm{d}z=$$

$$\int_{-\infty}^{\infty}v_0\sum_{l=0}^{\lambda+1}C_l^{\lambda+1}q^l\exp\left(\frac{l(2z\Delta s-\Delta s^2)}{2\sigma_i^2}\right)(1-q)^{\lambda+1-l}\mathrm{d}z=$$

$$\sum_{l=0}^{\lambda+1}C_l^{\lambda+1}q^l\exp\left(\frac{l(l-1)\Delta s^2}{2\sigma_i^2}\right)(1-q)^{\lambda+1-l}\leqslant$$

$$\left(q\exp\left(\frac{\lambda\Delta s^2}{2\sigma_i^2}\right)+1-q\right)^{\lambda+1}\leqslant$$

$$\exp\left(q(\lambda+1)\left(\exp\left(\frac{\lambda\Delta s^2}{2\sigma_i^2}\right)-1\right)\right) \tag{21}$$

当 $\lambda\in[1,T]$ 时，有

$$E_{v_1,v_0}\leqslant\exp(q(\lambda+1)\left(\frac{l(2z\Delta s-\Delta s^2)}{2\sigma_i^2}+O\left(\frac{\lambda^2\Delta s^4}{4\sigma_i^4}\right)\right)\approx$$

$$\exp\left(\frac{q\lambda(\lambda+1)\Delta s^2}{2\sigma_i^2}\right) \tag{22}$$

将该结果代入时刻损失函数 $\alpha(\lambda)$ 中，可得

$$\alpha(\lambda)\leqslant\sum_{t=1}^{T}\alpha(\lambda,\sigma_i)=\frac{Tq\lambda(\lambda+1)\Delta s^2}{2\sigma_i^2} \tag{23}$$

根据定义 2 中的尾约束可知，当式(24)成立时，添加噪声标准差为 σ_i 的隐私机制可以满足 (ε_i,δ_i) – 差分隐私。

$$\delta=\min_{\lambda}\exp\left(\frac{Tq\lambda(\lambda+1)\Delta s^2}{2\sigma_i^2}-\lambda\varepsilon_i\right) \tag{24}$$

结合式(23)，可得

$$\frac{Tq\lambda(\lambda+1)\Delta s^2}{2\sigma_i^2}-\lambda\varepsilon_i=$$

$$\frac{Tq\Delta s^2}{2\sigma_i^2}\left(\lambda+\frac{1}{2}-\frac{\varepsilon_i\sigma_i^2}{Tq\Delta s^2}\right)^2-\frac{Tq\Delta s^2}{2\sigma_i^2}\left(\frac{1}{2}-\frac{\varepsilon_i\sigma_i^2}{Tq\Delta s^2}\right)^2 \tag{25}$$

令 $\lambda=\dfrac{\varepsilon_i\sigma_i^2}{Tq\Delta s^2}-\dfrac{1}{2}$，可得

$$\ln\left(\frac{1}{\delta}\right)\leqslant\frac{Tq\Delta s^2}{2\sigma_i^2}\left(\frac{1}{2}-\frac{\varepsilon_i\sigma_i^2}{Tq\Delta s^2}\right)^2=$$

$$\frac{Tq\Delta s^2}{8\sigma_i^2}-\frac{\varepsilon_i}{2}+\frac{\varepsilon_i^2\sigma_i^2}{2Tq\Delta s^2} \tag{26}$$

由于 $\delta_i\in(0,1)$，可得

$$\frac{Tq\lambda(\lambda+1)\Delta s^2}{2\sigma_i^2}-\lambda\varepsilon_i<0 \tag{27}$$

结合式(24)，可以对 $\ln\left(\dfrac{1}{\delta}\right)$ 做如式(28)所示的约束

$$\ln\left(\frac{1}{\delta}\right)<-\frac{\varepsilon_i}{4}+\frac{\varepsilon_i\sigma_i^2}{2Tq\Delta s^2}<\frac{\varepsilon_i\sigma_i^2}{2Tq\Delta s^2} \tag{28}$$

因此，选择 $\sigma_i=\dfrac{\Delta s\sqrt{2qT\ln\left(\dfrac{1}{\delta_i}\right)}}{\varepsilon_i}$ 时，联邦学习中每个参与方的本地训练过程满足 (ε_i,δ_i) – 本地化差分隐私，从而保证参与方的本地训练数据不会被诚实且好奇的其他参与方或中心服务器以及外部攻击者所窃取。

3.2　算法复杂度分析

本节分析算法的时间复杂度，LDP-FL 方法由 FL_Server 和 FL_Client 这 2 个算法组成，FL_Client 算法嵌套在 FL_Server 中，记 LDP-FL 方法的整体迭代次数为 T，参与方数量为 N，在每次迭代时，FL_Client 算法的时间复杂度为 $O(\log(N))$，则 LDP-FL 方法的时间复杂度等于 FL_Server 算法的时间复杂度，为 $O(T\log(N))$。

4 实验与分析

4.1 实验设置

4.1.1 实验环境

本节对本文所提 LDP-FL 方法的有效性进行评估，并设计对比实验。所使用的实验平台操作系统为 Windows 10(64 位)，开发环境为 Pycharm，编程语言为 Python 3.8，CPU 为 11th Gen Intel(R) Core(TM) i5-11400H @ 2.70 GHz，内存为 16 GB。实验使用 Pytorch1.7.1 训练深度学习模型，采用卷积神经网络（CNN, conventional neural network）构建本文所提 LDP-FL 方法，设置 2 个卷积层分别有 16 和 32 个特征，并使用一个 5×5、步长为 2 的卷积核，以及一个输入张量为 7×7×32、输出张量为 10 的全连接层，采用梯度下降进行模型训练时所选择的批次大小为 64，参与方本地训练迭代次数为 10 次。

4.1.2 实验数据集

实验采用 2 种数据集，分别是 MNIST 数据集和 Fashion MNIST 服饰数据集。其中，MNIST 数据集包含 10 种手写数字识别的灰度图像数据，有 60 000 个训练图像和 10 000 个测试图像，每个灰度图像包含 28 像素×28 像素；Fashion MNIST 服饰数据集是经典 MNIST 数据集的简易替换，比常规 MNIST 手写数据将更具挑战性，包含 60 000 个示例的训练集和 10 000 个示例的测试集，每个示例都是一个 28 像素×28 像素灰度图像，可以分为 10 种类型。

4.1.3 评价指标

为了验证本文所提 LDP-FL 方法的优越性，选择原始的联邦平均方法 FedAvg[1]作为参照，并将 LDP-FL 方法与 CL-FL 方法[12]和 NbAFL 方法[15]进行对比，主要的评价指标有以下 3 种。

1) 全局准确率。经过多次迭代后，联邦模型的全局准确率是衡量算法有效性的关键指标。通过对比相同条件下不同算法的全局准确率，可以直观地判断算法的性能。

2) 性能损失。性能损失是衡量联邦模型性能的指标，通过性能估计机制进行计算。

3) 运行时间。算法的运行时间是衡量通信开销的重要指标。运行时间越长，则通信开销越大。

4.2 有效性衡量实验

本节探究 LDP-FL 有效性。使用 MNIST 数据集，联邦学习迭代轮次 $T=150$，设置 $\delta=0.001$，每个参与方的隐私预算 $\varepsilon_i=\varepsilon$，$\sigma_i=10^{-5}$。首先，探究隐私预算对全局准确率的影响，在采样率 $q=1$、参与方 $N=10$ 的情况下，分别设置隐私预算 $\varepsilon=1.0,\varepsilon=2.0,\varepsilon=4.0$，结果如图 2(a)所示。其次，探究参与方数量的影响，在采样率 $q=1$、隐私预算 $\varepsilon=1.0$ 的情况下，分别设置参与方数量为 $N=10,N=50,N=100$，结果如图 2(b)所示。

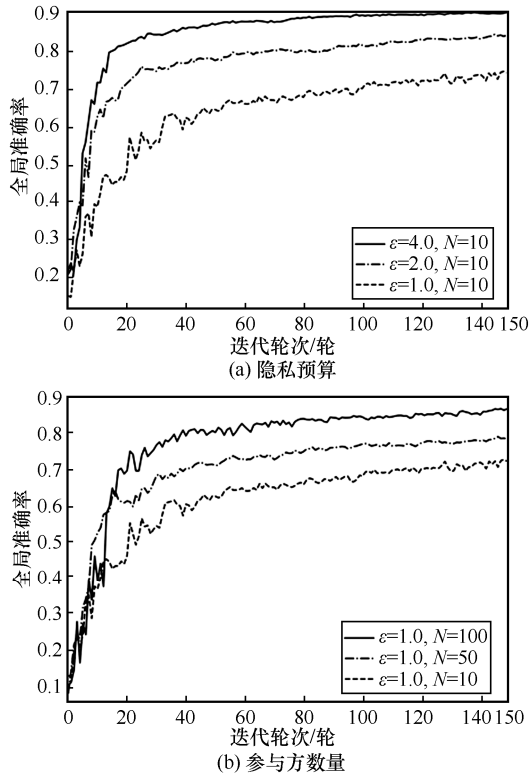


图 2 LDP-FL 方法有效性衡量实验

观察图 2，可以得到如下结论。

1) 在参与方数量和采样率均相同的前提下，LDP-FL 方法中隐私预算越高，模型全局准确率越高，说明可以通过调整隐私预算实现联邦学习模型隐私性和可用性的平衡。

2) 在隐私预算和采样率均相同的前提下，LDP-FL 方法中参与方数量越多，模型全局准确率越高，说明增加联邦学习参与方数量可以提高准确率。

3) 在以上实验中，经过大约 80 次迭代后，LDP-FL 方法的全局准确率趋于稳定，说明模型可用性较好。

4.3 对比实验与分析

4.3.1 全局准确率对比

首先，探究本文所提 LDP-FL 方法与现有方

法在 MNIST 数据集和 Fashion MNIST 数据集上全局准确率的对比情况。对于 4 种联邦学习方法，设置参与方数量 $N=10$，采样率 $q=1$，对于使用差分隐私的 LDP-FL、CL-FL 和 NbAFL，设置每个参与方的隐私预算 $\varepsilon_i=\varepsilon$，$\sigma_i=10^{-5}$，总体隐私预算 $\varepsilon=4.0$，图 3 分别展示了 4 种联邦学习方法在 2 种数据集上的全局准确率随迭代轮次的变化情况。

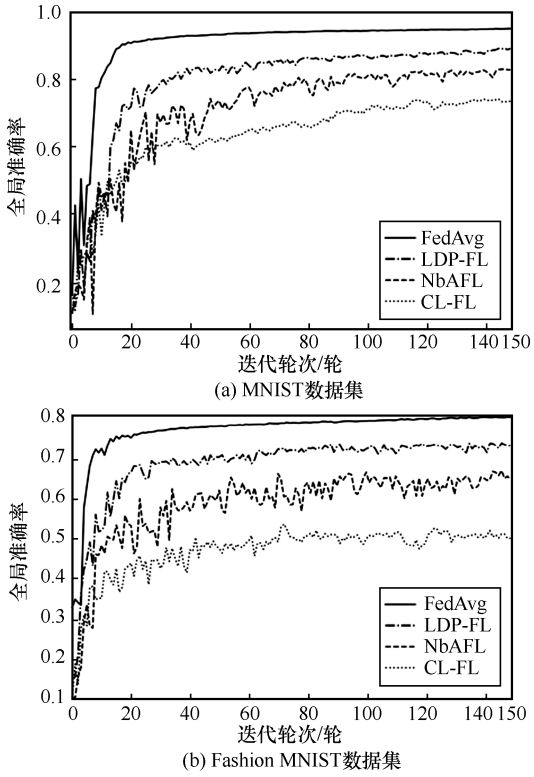


图 3　全局准确率随迭代轮次的变化情况

观察图 3，可以得到如下结论。

1) 在参与方数量相同的情况下，引入差分隐私保护的 LDP-FL 方法、CL-FL 方法和 NbAFL 方法的全局准确率在 2 种数据集上均低于 FedAvg，说明与 FedAvg 相比，引入噪声机制会对联邦学习模型的准确率造成影响。

2) 在参与方数量和隐私预算均相同的情况下，本文所提的 LDP-FL 方法全局准确率在 2 种数据集上均高于 CL-FL 方法和 NbAFL 方法，说明 LDP-FL 方法的性能优于 CL-FL 方法和 NbAFL 方法。

3) 由于 Fashion MNIST 数据集中的图像数据比 MNIST 数据集中数据更复杂，因此 4 种方案在 MNIST 数据集上的表现均优于在 Fashion MNIST 数据集上的表现。

4.3.2　性能损失对比

其次，探究本文所提 LDP-FL 方法与 CL-FL 方法和 NbAFL 方法在 MNIST 数据集和 Fashion MNIST 数据集上性能损失的对比情况。设置参与方数量 $N=10$，采样率 $q=1$，每个参与方的隐私预算 $\varepsilon_i=\varepsilon$，$\sigma_i=10^{-5}$，总体隐私预算 $\varepsilon=4.0$，表 3 分别展示了 3 种联邦学习方案在 2 种数据集上性能损失对比实验结果。

通过表 3 可以看出，LDP-FL 方法在 2 种数据集上不同的迭代轮次下的性能损失值均小于 CL-FL 方法和 NbAFL 方法，说明 LDP-FL 方法的性能优于 2 种对比算法。

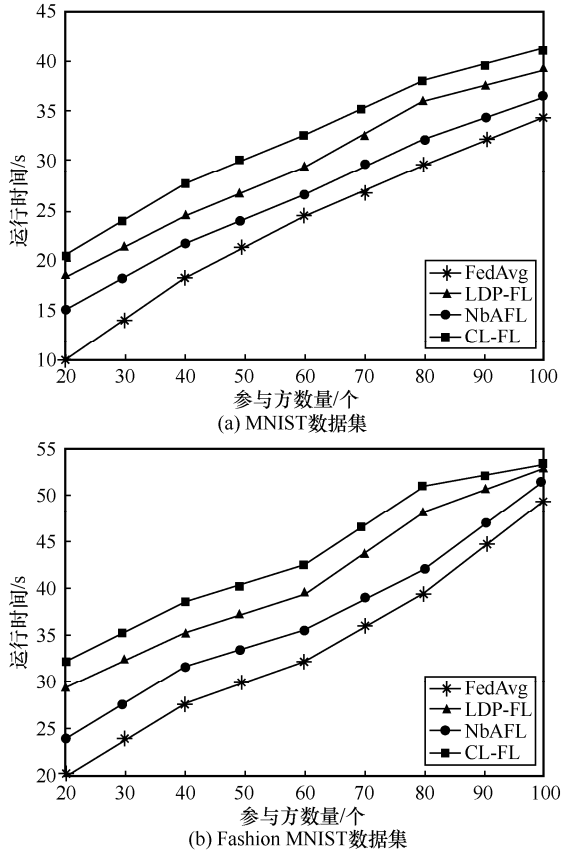
4.3.3　算法运行时间对比

最后，探究本文所提 LDP-FL 方法与现有方法在 MNIST 数据集和 Fashion MNIST 数据集上运行时间的对比情况。对于 4 种联邦学习方法，分别设定参与方数量为 $N=[20,40,60,80]$，对于使用差分隐私的 LDP-FL、CL-FL 和 NbAFL，设

表 3　　**性能损失对比实验结果**

数据集	方法	迭代轮次/次					
		25	50	75	100	125	150
MNIST	LDP-FL	1.87	1.84	1.82	1.72	1.75	1.72
	NbAFL	1.89	1.88	1.90	1.81	1.82	1.79
	CL-FL	1.92	1.90	1.99	1.84	1.88	1.85
Fashion MINST	LDP-FL	2.21	2.14	2.13	2.20	2.17	2.15
	NbAFL	2.33	2.24	2.15	2.44	2.35	2.46
	CL-FL	2.52	2.44	2.47	2.58	2.68	2.77

置每个参与方的隐私预算 $\varepsilon_i = \varepsilon$，$\sigma_i = 10^{-5}$，总体隐私预算 $\varepsilon = 4.0$。图 4 分别展示了 4 种联邦学习方法在 2 种数据集上运行时间随参与方数量的变化情况。



(a) MNIST数据集

(b) Fashion MNIST数据集

图 4 运行时间随参与方数量的变化情况

观察图 4，可以得到如下结论。

1) 随着参与方数量的增加，4 种方法在 2 个数据集上的运行时间均有所增加，说明增加参与方数量会导致算法运行时间增加。

2) 由于 Fashion MNIST 数据集中的图像数据比 MNIST 数据集中数据更复杂，因此 4 种方法在 MNIST 数据集上的运行时间比 Fashion MNIST 数据集上的运行时间短。

3) 在参与方数量相同的情况下，FedAvg 方法的运行时间最短；在 3 种引入噪声机制的联邦学习隐私保护方案中，NbAFL 方法的运行时间最短，本文所提 LDP-FL 方法略次之，CL-FL 方法最长，同样说明了 LDP-FL 方法的有效性。

5 结束语

本文通过设计一种基于本地化差分隐私的联邦学习方法 LDP-FL，解决联邦学习中存在的模型推理攻击，主要是将该机制作用在联邦学习参数的传递过程中，增加联邦模型训练的隐私性。同时，设计一种适用于联邦学习的性能损失约束机制，通过优化损失函数的约束范围来降低本地化差分隐私联邦模型的性能损失。最后在真实的数据集上通过实验验证了所提 LDP-FL 方法的有效性。未来的工作将集中在联邦学习优化，以及隐私保护联邦学习在应用方面的拓展，如医疗和物联网环境，研究这些场景下如何在保证隐私安全的同时提高联邦模型的全局准确率。

参考文献：

[1] YANG Q, LIU Y, CHEN T J, et al. Federated machine learning: concept and applications[J]. ACM Transactions on Intelligent Systems and Technology, 2019, 10(2): 1-19.

[2] KAIROUZ P, MCMAHAN H B, AVENT B, et al. Advances and open problems in federated learning[J]. Foundations and Trends in Machine Learning, 2021, 14(1-2): 1-210.

[3] 方晨, 郭渊博, 王一丰, 等. 基于区块链和联邦学习的边缘计算隐私保护方法[J]. 通信学报, 2021, 42(11): 28-40.
FANG C, GUO Y B, WANG Y F, et al. Edge computing privacy protection method based on blockchain and federated learning[J]. Journal on Communications, 2021, 42(11): 28-40.

[4] 莫梓嘉, 高志鹏, 杨杨, 等. 面向车联网数据隐私保护的高效分布式模型共享策略[J]. 通信学报, 2022, 43(4): 83-94.
MO Z J, GAO Z P, YANG Y, et al. Efficient distributed model sharing strategy for data privacy protection in Internet of vehicles[J]. Journal on Communications, 2022, 43(4): 83-94.

[5] 刘艺璇, 陈红, 刘宇涵, 等. 联邦学习中的隐私保护技术[J]. 软件学报, 2022, 33(3): 1057-1092.
LIU Y X, CHEN H, LIU Y H, et al. Privacy-preserving techniques in federated learning[J]. Journal of Software, 2022, 33(3): 1057-1092.

[6] SONG M K, WANG Z B, ZHANG Z F, et al. Analyzing user-level privacy attack against federated learning[J]. IEEE Journal on Selected Areas in Communications, 2020, 38(10): 2430-2444.

[7] LIU X Y, LI H W, XU G W, et al. Privacy-enhanced federated learning against poisoning adversaries[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 4574-4588.

[8] PHONG L T, AONO Y, HAYASHI T, et al. Privacy-preserving deep learning via additively homomorphic encryption[C]//Proceedings of IEEE Transactions on Information Forensics and Security. Piscataway: IEEE Press, 2019: 1333-1345.

[9] OU W, ZENG J, GUO Z, et al. A homomor-phic-encryption-based vertical federated learning scheme for rick management[J]. Computer Science and Information Systems, 2020, 17(3): 819-834.

[10] ZHU H F, MONG G R S, NG W K. Privacy-preserving weighted

federated learning within the secret sharing framework[J]. IEEE Access, 2020, 8: 198275-198284.

[11] DWORK C. Differential privacy[C]//Proceedings of 2006 International Colloquium on Automata, Languages and Programming (ICALP). Berlin: Springer, 2006: 1-12.

[12] GEYER R C, KLEIN T, NABI M. Differentially private federated learning: a client level perspective[J]. arXiv Preprint, arXiv: 1712.07557, 2017.

[13] 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 Press, 2016: 308-318.

[14] ZHAO C X, SVN Y, WANG D G. Federated learning with Gaussian differential privacy[C]//Proceedings of the 2020 2nd International Conference on Robotics, Intelligent Control and Artificial Intelligence. Piscataway: IEEE Press, 2020: 296-301.

[15] WEI K, LI J, DING M, et al. Federated learning with differential privacy: algorithms and performance analysis[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 3454-3469.

[16] TRUEX S, LIU L, CHOW K H, et al. LDP-Fed: federated learning with local differential privacy[C]//Proceedings of the Third ACM International Workshop on Edge Systems, Analytics and Networking. New York: ACM Press, 2020: 61-66.

[17] LIU R X, CAO Y, YOSHIKAWA M, et al. FedSel: federated SGD under local differential privacy with top-k dimension selection[C]//International Conference on Database Systems for Advanced Applications. Berlin: Springer, 2020: 485-501.

[18] ZHAO Y, ZHAO J, YANG M M, et al. Local differential privacy-based federated learning for Internet of things[J]. IEEE Internet of Things Journal, 2021, 8(11): 8836-8853.

[19] MCMAHAN H B, MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data[J]. arXiv Preprint, arXiv: 1602.05629, 2016.

[作者简介]

康海燕（1971− ），男，河北灵寿人，博士，北京信息科技大学教授，主要研究方向为网络安全与隐私保护等。

冀源蕊（1997− ），女，宁夏银川人，北京信息科技大学硕士生，主要研究方向为网络安全与隐私保护。