计算机科学
COMPUTER SCIENCE

参数解耦在差分隐私保护下的联邦学习中的应用

王梓行, 杨敏, 魏子重

引用本文

王梓行, 杨敏, 魏子重. 参数解耦在差分隐私保护下的联邦学习中的应用[J]. 计算机科学, 2024, 51(11): 379-388.

WANG Zihang, YANG Min, WEI Zichong. Application of Parameter Decoupling in Differentially Privacy Protection Federated Learning [J]. Computer Science, 2024, 51(11): 379-388.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.231200034

参数解耦在差分隐私保护下的联邦学习中的应用

王梓行[1]　杨　敏[1]　魏子重[2]
1 空天信息安全与可信计算教育部重点实验室(武汉大学国家网络安全学院)　武汉 430072
2 浪潮集团科学研究院　济南 250101
(wzihang@whu.edu.cn)

摘　要　联邦学习(Federated Learning,FL)是一种先进的隐私保护机器学习技术,其通过多方协作,在无需集中聚合原始数据的情况下,交换模型参数以训练共享模型。尽管在FL中参与方不需要显式地共享数据,但许多研究表明,其仍然面临多种隐私推理攻击,从而导致隐私信息泄露。为应对这一问题,学术界提出了多种解决方案。其中,一种严格保障隐私的方法是将本地化差分隐私(Local Differential Privacy,LDP)技术应用于联邦学习。该技术在参与方上传模型参数前对其添加随机噪声,能有效地抵御恶意攻击者的推理攻击。然而,LDP引入的噪声会造成模型性能下降。同时,最新研究指出,这种性能下降与LDP在客户端之间引入了额外的异构性有关。针对LDP使得FL性能下降的问题,提出了差分隐私保护下基于参数解耦的联邦学习方案(PD-LDPFL):除了服务器下发的基础模型外,每个客户端在本地还额外学习了个性化输入和输出模型。该方案在客户端传输时仅上传添加噪声后的基础模型的参数,而个性化模型的参数被保留在本地,自适应改变客户端本地数据的输入和输出分布,缓解LDP引入的额外异构性以减少精度损失。此外,研究发现,即使在采用较高的隐私预算的情况下,该方案也能天然地抵御一些基于梯度的隐私推理攻击,如深度梯度泄露等攻击方法。在MNIST,FMNIST和CIFAR-10这3个常用数据集上进行了实验,结果表明:相比传统的差分隐私联邦学习方法,该方案不仅可以获得更好的性能,而且还提供了额外的安全性。

关键词:联邦学习;差分隐私;异构性;参数解耦;隐私保护
中图分类号　TP309

Application of Parameter Decoupling in Differentially Privacy Protection Federated Learning

WANG Zihang[1],YANG Min[1] and WEI Zichong[2]
1 Key Laboratory of Aerospace Information Security and Trusted Computing,Ministry of Education,School of Cyber Science and Engineering,Wuhan University,Wuhan 430072,China
2 Inspur Group Scientific Research Institute,Jinan 250101,China

Abstract　Federated learning(FL) is an advanced privacy preserving machine learning technique that exchanges model parameters to train shared models through multi-party collaboration without the need for centralized aggregation of raw data. Although participants in FL do not need to explicitly share data,many studies show that they still face various privacy inference attacks,leading to privacy information leakage. To address this issue,the academic community has proposed various solutions. One of the strict privacy protection methods is to apply Local differential privacy(LDP) technology to federated learning. This technology adds random noise to the model parameters before they are uploaded by participants,to effectively resist inference attacks from malicious attackers. However,the noise introduced by LDP can reduce the model performance. Meanwhile,the latest research suggests that this performance decline is related to the additional heterogeneity introduced by LDP between clients. A parameter decoupling based federated learning scheme(PD-LDPFL) with differential privacy protection is proposed to address the issue of FL performance degradation caused by LDP. In addition to the basic model issued by the server,each client also learns personalized input and output models locally. This scheme only uploads the parameters of the basic model with added noise during client transmission,while the personalized model is retained locally,adaptively changing the input and output distribution of the client's local data to alleviate the additional heterogeneity introduced by LDP and reduce accuracy loss. In addition,research has found that even with a higher privacy budget,this scheme can naturally resist some gradient based privacy inference attacks,such as deep gradient leakage and other attack methods. Through experiments on three commonly used datasets,MNIST,FMNIST,and CI-

到稿日期:2023-12-05　返修日期:2024-04-02
基金项目:国家自然科学基金(62172308);国家重点基础研究发展计划(2021YFB2700200)
This work was supported by the National Natural Science Foundation of China(62172308) and National Basic Research Program of China(2021YFB2700200).
通信作者:杨敏(yangm@whu.edu.cn)

FAR-10, the results show that this scheme not only achieves better performance compared to traditional differential privacy federated learning, but also provides additional security.

Keywords Federated learning, Differential privacy, Heterogeneity, Parameter decoupling, Privacy preserving

1 引言

近年来，大数据驱动的人工智能技术在无人驾驶[1]、图像识别[2]、智能医疗[3]等领域得到了广泛应用。这些人工智能模型的训练通常需要大量的数据，但实际情况是，大部分有价值的数据分散在不同的组织间。由于法律限制和数据隐私安全的考虑，这些数据不能被直接共享。为解决这一问题，联邦学习[4]被提出。FL 作为一种具有隐私保护的分布式机器学习技术，允许各方通过上传模型参数的方式进行协同训练，从而创建一个共享的模型。这种分布式训练方式可以避免数据直接共享带来的隐私风险，同时充分利用各方的数据资源。然而，最新研究表明[5-6]，FL 在防范已知的隐私推理攻击方面仍存在一些缺陷。这是因为模型的参数是在各方本地数据集上训练得到的，不可避免地会携带本地数据集的信息，因此攻击者可以通过分析模型权重来获取客户端的敏感数据。例如，Aono 等[7]表明在联邦学习系统中从梯度中恢复图像是可行的；Yin 等[8]进一步在基于梯度反演的过程中引入保真正则化(FR)和组一致性正则化(GCR)，使攻击者能够从深度神经网络中通过 ResNet 恢复单个数据点(ImageNet50 级)；Geiping 等[9]发现，恶意服务器可以通过梯度反演攻击来恢复客户端的训练数据。因此，为确保安全性，FL 通常需要引入额外的隐私保护方法[10]。

差分隐私[11-12](Differential Privacy，DP)是一种安全可证明的隐私保护方法，从结构上看，其可分为中心化差分隐私和本地化差分隐私(Local Differential Privacy，LDP)[13]。在 FL 中，客户端可以在上传模型参数前向其添加满足 LDP 的噪声来保护隐私。这种方法使得即使攻击者获得扰动后的模型参数，也难以进行隐私推理攻击。值得注意的是，在 FL 中应用 LDP 至少还存在两个挑战。首先，LDP 向模型的参数中添加了随机噪声，尽管在聚合阶段噪声可以被平均，但仍不可避免地会对模型的精度产生影响，这在一些高精度任务中是不可接受的[14-16]。其次，在实际场景中，各方的数据通常是非独立同分布的[17]，这种异构性会导致模型的收敛性变差，而 LDP 添加的噪声会使模型更新方向产生偏离，从而在客户端之间引入额外的异构性，进一步降低模型的收敛性能。因此，最新的研究日益聚焦于从 LDP 算法本身或者模型结构的角度来减小由差分隐私带来的精度损失。例如，Fu 等[18]提出了 Adap DP-FL，通过添加自适应噪声有效提高了联邦学习中模型的精度。Huang 等[19]针对客户端间的非平衡数据提出了 DP-FL 框架，该框架根据每个客户端的数据量分配不同的隐私预算，以更新每个用户的模型参数。Yang 等[20]提出了 Privatefl，通过引入个性化输入转换来平衡差分隐私造成的精度损失。其中，个性化输入转换可以看作一个转换函数，对神经网络的输入进行变换，其参数通过本地训练得到。然而，我们发现个性化输入转换的参数学习需要经过一个过程。例如，在 50 轮的联邦学习迭代中，使用 Privatefl 方法训练的全局模型在测试集上的前 10 轮的准确率提升速度与传统差分隐私联邦学习方法相比并不显著。这表明本地客户端在早期迭代时仍会受到 LDP 引入的异构性的影响，不可避免地造成训练结果的偏差。

为解决上述问题，本文提出了一种基于参数解耦的差分隐私联邦学习方案 PD-LDPFL。该方案可以有效减小 LDP 引入的异构性，并且天然抵御常见的隐私推理攻击。如图 1 所示，在 PD-LDPFL 中，每个客户端的本地模型由个性化输入模型、神经网络基础模型和个性化输出模型组成，形成了一个拓展模型。

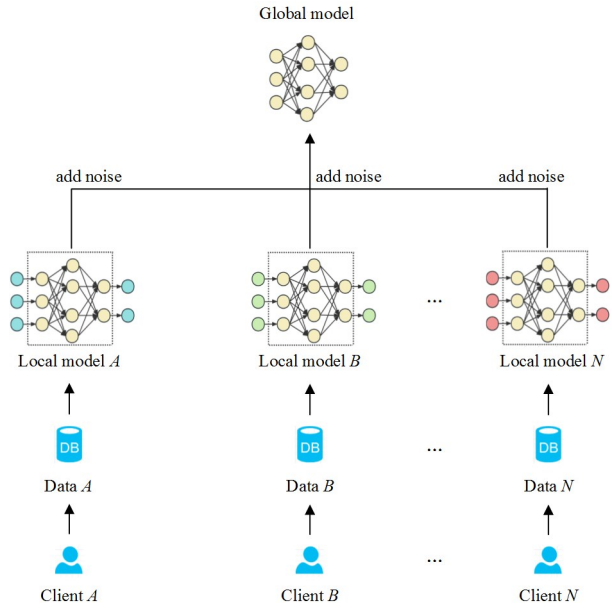


图 1 参数解耦的示例图

Fig. 1 Illustration of parameter decoupling

具体而言，首先考虑到每个客户端向基础模型的参数添加噪声会引入额外的异构性，我们设置个性化输入模型，用于自动学习数据的转换，从而改变客户端数据的原始分布，以补偿 LDP 引入的异构性。其次，由于个性化输入模型的参数学习需要一定的过程，并且添加的噪声会影响聚合后的输出准确率，因此我们同时引入了个性化输出模型。这一模型旨在自适应地学习基础模型的输出分布，以加快模型收敛的速度，同时使得每个客户端拥有独特的拓展模型。接着，使用理论分析和实验证明了参数解耦天然抵御常见隐私推理攻击的能力。然后通过进一步实验，分析了个性化模型的设置，包括变换函数选择和参数初始化，并给出了个性化模型设置的建议。其中的主要挑战在于如何找到尽可能最优的个性化输入与输出模型的设置，以减少客户端间的异构性。最后，在 3 个常用的图像分类数据集 MNIST，FMNIST 和 CIFAR-10 上进行了大量实验，评估了 PD-LDPFL 方案在常见因素影响下的性能。实验结果表明，在相同的数据分布和隐私保护程度下，PD-LDPFL 的性能显著优于最基础的 Fedavg 和相关方案

Privatefl。本文的主要贡献包括3个方面：

1）提出了基于参数解耦的差分隐私联邦学习方案，它在每个客户端本地额外学习个性化输入与输出模型，减少了LDP引入的异构性，并且能够天然抵御常见的隐私推理攻击。

2）深入分析了PD-LDPFL抵御隐私推理攻击和减小LDP引入的异构性的原因，并通过实验给出了个性化模型设置的建议。

3）在常用图像分类数据集上的实验证明了PD-LDPFL的优越性，其为解决联邦学习中的隐私泄露问题和异构性问题提供了一种有效的解决方案。

2 相关技术

本章将介绍有关FL和LDP的初步知识，以及其他必要的背景知识。此外，我们还将从客户端间损失函数偏差的角度介绍LDP引入的异构性。

2.1 联邦学习

联邦学习的概念最早在2016年由谷歌提出，最初旨在解决安卓手机终端用户在本地更新模型的问题[21-22]。其设计目标是在保障大数据交换时的信息安全、保护终端数据和个人数据隐私的前提下，在多参与方或多计算节点之间进行高效率的人工智能模型训练。一般而言，联邦学习包括以下步骤：首先，中心服务器将初始化的全局模型θ_0下发到每个客户端；其次，在第r轮，客户端k利用本地数据训练出本地模型θ_k^r，并将模型的参数上传到服务器；最后，服务器将每个客户端的参数聚合，得到一个新的全局模型，即$\theta^{r+1}=\frac{1}{K}\sum_{k=1}^{K}\theta_k^r$，其中$K$是参与联邦训练的客户端总数，$r$是训练的轮次。上述步骤重复执行，直到模型收敛或者训练次数达到一定轮次。

2.2 本地化差分隐私

差分隐私是隐私保护数据收集领域的最先进技术之一。传统的差分隐私需要一个可信的第三方服务器，然而在实际应用中很难实现这一点。为了消除这一限制，本地化差分隐私[23-24]被提出并被应用于联邦学习，以对抗FL中普遍存在的隐私推理攻击。在FL中，当每个客户端训练完其本地模型后，可以使用LDP算法φ来扰动其模型参数θ，并上传$\varphi(\theta)$从而使其免受多种推理攻击。设D表示数据集，LDP的定义如定义1所示。

定义1（ϵ-本地化差分隐私） 一个随机化算法M满足ϵ-本地化差分隐私（$\epsilon>0$），当且仅当数据集D中的任意一对输入值v_1，v_2，有：

$$\forall y\in Range(M):\frac{\Pr[M(v_1)=y]}{\Pr[M(v_2)=y]}\leqslant e^{\epsilon} \tag{1}$$

其中，$Range(M)$代表算法M的所有可能输出。参数是隐私预算，它量化了隐私保护的水平，通常被置为一个相对较小的值。当越小时，隐私保护程度越强，反之则隐私保护程度越弱。

由于FL中模型的参数为数值型数据，因此可以采用适用于数值型数据的本地化差分隐私机制。这些机制包括拉普拉斯机制[25]、高斯机制[26]以及随机响应机制[27]等。其中，高斯机制使用最广泛，它不仅可用于模型参数的隐私保护，还适用于样本级别的隐私保护[28]（即在训练步骤中进行梯度裁剪并添加噪声），并能通过矩会计[29]方法准确追踪隐私损耗。在本文的场景中，针对训练后的模型参数，我们选择了随机响应机制中的分段机制算法（Piecewise Mechanism，PM）[30]。该算法在相同隐私预算下的方差低于其他算法。

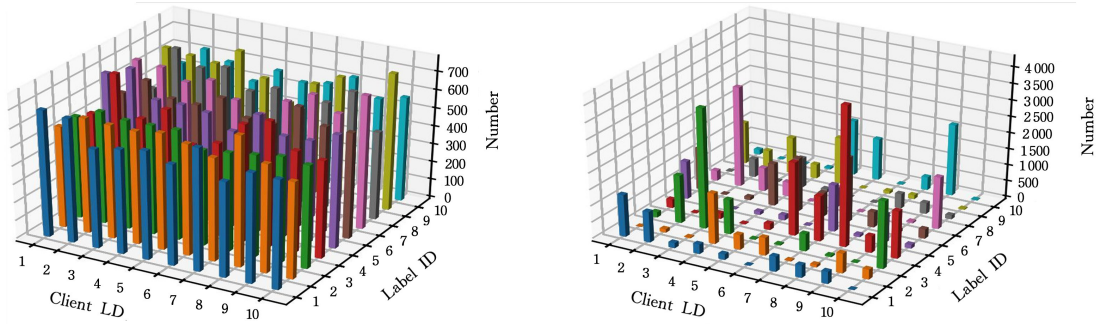
2.3 隐私推理攻击

隐私推理攻击指攻击者试图从联邦学习模型的更新或输出中推断出有关客户端数据的敏感信息。这些攻击通常利用模型参数、梯度或输出的统计信息，来推断关于训练数据的隐私信息，例如个人身份、特定属性或敏感特征等。在众多隐私推理攻击中，深度梯度泄露（Deep Leakage from Gradients，DLG）[31]是一种常见且实施简单的攻击方法。例如，在图像分类任务的联邦学习中，即使参与方的数据并未直接共享，这种攻击技术也可以使得攻击者能够推断出参与方的图像数据。其基本原理是攻击者或诚实且好奇的服务器可以生成与参与方训练集图片分辨率相同的随机图片。随后，他们用随机图片训练模型，将得到的梯度和参与方用隐私图片训练出来的梯度进行比较，并通过平方损失得到一个距离的差值。然后攻击者持续添加扰动，以减少这个差值。最后，当差值收敛时，攻击就完成了，因为攻击者图片和参与方图片在这个模型中的梯度几乎是一致的。

2.4 本地化差分隐私引入的异构性

联邦学习通常面临客户端之间的数据分布[32-34]、模型结构[35-36]、网络环境[37-38]和硬件设备[39]的异构性影响。其中，客户端间数据分布的异构性对联邦学习性能的影响最为显著。

如图2所示，当数据是独立同分布（Independent Identically Distribution，I. I. D）时，每个客户端拥有的样本数量与对应的标签是非常接近的；而当数据是非独立同分布（Non-Independent Identically Distribution，Non-I. I. D）时，客户端之间的样本数量与标签类别都存在较大差异。由于客户端之间不能直接共享数据，其数据分布是不同的，那么协同训练得到的模型就是有偏的，并且性能通常较差。而将差分隐私应用于训练后的模型参数，向模型参数添加随机噪声，会进一步使得模型的更新方向偏离原本的优化方向，从而增加了客户端之间的数据异构性。假设联邦学习训练总轮次为R，客户端总数量为K，第r轮中客户端k在其本地数据集训练后的损失函数值为$l(X_k;Y_k)^r$，则这种异构性可以由客户端间损失函数值的R轮累计方差来衡量，即$l_{\text{sum}}=\frac{1}{K}\sum_{r=1}^{R}\sum_{k=1}^{K}(l(X_k;Y_k)^r-l(\overline{X;Y})^r)^2$，其中$l(\overline{X;Y})^r=\frac{1}{K}\sum_{k=1}^{K}l(X_k;Y_k)^r$，$X_k$是客户端$k$的本地训练集，$Y_k$是其标签。根据上述计算式，在每一轮联邦学习训练中，可以计算训练后参与方之间损失函数值的方差，这种方差在理想情况下是一个很小的值。即当参与方的本地训练数据为独立同分布时，如果不使用LDP，那么训练后参与方之间的损失函数值应该是非常接近的。然后，计算总轮次中参与方损失函数值的累计方差。在相同条件下，当加入LDP噪声后，累计方差值越大，说明LDP引入的异构性越强。



(a) I. I. D (b) Non-I. I. D

图 2 样本在客户端间分布的 3D 柱状图

Fig. 2 3D bar chart of sample distribution between clients

3 基于参数解耦的差分隐私联邦学习方案

本章首先介绍了个性化模型的概念和添加个性化模型后的客户端的优化目标，然后提出了基于参数解耦的差分隐私联邦学习方案 PD-LDPFL。

3.1 个性化模型

作为客户端本地的保留模型，个性化模型的本质是一个保留在本地的转换函数，它能够自适应地调整客户端模型的输入或输出。由于其参数是通过训练学习得到的，因此我们称之为个性化模型。假设客户端 k 的本地数据集为 X_k，预测值为 $\hat{Y}_k$，θ_k^{Ti} 和 θ_k^{To} 分别表示第 k 个客户端的个性化输入模型和个性化输出模型。那么客户端 k 的输入经过个性化输入模型变换为 $\theta_k^{Ti}(X_k)$，输出经过个性化输出模型变换为 $\theta_k^{To}(\hat{Y}_k)$。其中 θ_k^{Ti} 用于自主学习改变客户端的数据特征分布，θ_k^{To} 用于自适应改变客户端的输出分布。理论上，个性化模型的设置有着广泛的选择，但我们认为其需要满足特定要求。就个性化输入模型而言，它应该尽可能地保留数据的原始特征，以便让基础模型在后续训练中获取相关信息，同时也需要调整特征分布以应对异构性。而对于个性化输出模型，它应充分保留客户端本地的输出分布，并尽可能地适应本地数据集。

3.2 优化目标

在传统的 FL 中，每个客户端通过最小化损失函数来调整模型，以拟合标签和预测值之间的偏差[40]。假设 θ_k 表示客户端 k 的本地模型，D_k 表示客户端 k 拥有的数据集。对于客户端 k，其优化目标为找到一个 θ^*，使其满足：

$$\theta^* = \arg\min_{\theta} \frac{1}{|D_k|} \sum_{n=1}^{|D_k|} l(y^{(n)}, f(x^{(n)}; \theta_k)) \tag{2}$$

其中，$l(y^{(n)}, f(x^{(n)}; \theta))$ 是 D_k 中一个训练样本 $(x^{(n)}, y^{(n)})$ 对应于本地模型 θ_k 的损失。

考虑图 1 中的设置，每个客户端拥有独特的个性化输入和输出模型，并共享相同的基础模型。此时，客户端 k 的目标变为学习基础模型 θ_k 与个性化模型 θ_k^{Ti} 和 θ_k^{To}。因此，其优化目标变为找到一个 θ^*，使其满足：

$$\theta^* = \arg\min_{\theta} \frac{1}{|D_k|} \sum_{n=1}^{|D_k|} l(y^{(n)}, f(x^{(n)}; \theta_k^{Ti} + \theta_k + \theta_k^{To})) \tag{3}$$

其中，+ 表示基础模型与个性化模型的组合。具体而言，客户端 k 的每轮模型由基础模型以及个性化输入和输出模型构成。其中基础模型被初始化为服务器下发的全局模型 θ_k，个性化输入和输出模型被初始为上一轮学习后个性化输入和输出模型。

3.3 参数解耦方案的流程

3.2 节介绍了参数解耦方案中个性化模型的概念和拓展模型的优化目标，接下来我们将个性化模型的设置与 LDP 相结合，得到一个高效、个性化且具有隐私保护能力的联邦学习方案 PD-LDPFL。该方案由 3 个部分组成，分别是服务器聚合、客户端本地更新和 LDP 噪声添加。

3.3.1 PD-LDPFL 的整体流程

算法 1 描述了 PD-LDPFL 的整体流程，包括服务器和客户端分别要完成的工作。在算法开始时，服务器广播所需设置的隐私预算和其他必要参数，并将初始化后的全局模型 θ^0 发送给客户端。客户端完成本地训练后，服务器收集客户端基础模型的参数，并将其聚合得到新一轮的全局模型。客户端的优化目标是学习服务器下发的基础模型以及个性化输入和输出模型。算法 1 中的 ClientLocalUpdate 部分描述了这个过程。客户端 k 首先将本地模型初始化为服务器下发的基础模型，然后将基础模型与个性化输入模型和个性化输出模型组合得到一个拓展的本地模型。接下来，在每个本地训练轮次中，客户端 k 都使用随机梯度下降方法来优化拓展模型的权重，即端到端的训练过程为：客户端的数据先经过个性化输入模型处理后，再经过基础模型处理，最后由个性化输出模型输出。当训练完成后，拓展模型被分离为本地模型以及个性化输入和输出模型。个性化模型被保留在本地用于下一轮训练，而基础模型则被添加 LDP 噪声后上传到服务器。

算法 1 The workflow of PD-LDPFL

输入：客户端数量 K，所有参与方样本总数 n，第 k 个客户端样本数 n_k，训练轮次数 R，隐私预算 ϵ，学习率 γ，本地训练轮次 E 和本地小批量数据集 B

输出：最终模型 θ^R

1. 服务器执行：
2. 服务器初始化 θ^0 和 ϵ
3. for 每个训练轮次 r from 1 to R do
4. for 每个客户端 k from 1 to K 同时 do
5. $\theta_k \leftarrow$ ClientLocalUpdate(k, θ^{r-1}, ϵ)
6. $\theta^r \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} \theta_k$

7. 服务器将θ^r发送给各个客户端
8. ClientLocalUpdate(k,θ^{r-1},ϵ)
9. 客户端 k 初始化本地基础模型 $\theta_k^L \leftarrow \theta^{r-1}$ 和拓展模型 $\theta_k \leftarrow \theta_k^{Ti} + \theta_k^L + \theta_k^{To}$
10. for 本地的每轮迭代 i=1,2,…,E do
11. for 批量中每个样本和标签$(x_i, y_i) \in B$ do
12. for 每个批量 $b \in B$ do
13. $\theta_k = \theta_k - \gamma \nabla L(\theta_k; b)$
14. $\theta_k^L \leftarrow$将本地基础模型从拓展模型θ_k分离
15. $\theta_k^L \leftarrow$DataPertubation(θ_k^L,ϵ)
16. 客户端 k 将 θ_k^L发送给服务器

3.3.2 数据扰动

对模型参数进行扰动是 PD-LDPFL 中确保客户端隐私安全的关键步骤，这里采用了 LDP 机制中的 PM 算法。在 PM 中，输入在−1 到 1 之间，输出最多由 3 个“块”组成，其长度和位置取决于输入数据，最终输出被扰动为 3 个“块”之一的均匀采样值。当隐私预算越大时，PM 以更高概率在输入的附近采样。为了适应联邦学习环境，我们对 PM 算法进行了微小的改动。实现过程中，虽然神经网络的参数通常都是很小的值，满足 PM 算法对输入值的要求，但我们仍会先将其放大，这样做的好处是扰动后放缩可以进一步减小误差。具体的实现细节如算法 2 所示。

算法 2 DataPertubation

输入：本地模型 θ，隐私预算ϵ

输出：扰动后的本地模型θ^*

1. 初始化重要参数 $C \leftarrow \frac{e^{\epsilon/2}+1}{e^{\epsilon/2}-1}$，$t \leftarrow \max(\theta)$

/* $\max(\theta)$是模型 θ 的参数中最大的数值 */

2. $\theta \leftarrow \frac{\theta}{t}$
3. for 模型参数中的每个值 $v \in \theta$ do
4. $l(v) \leftarrow \frac{C+1}{2} \cdot v - \frac{C-1}{2}$
5. $r(v) \leftarrow l(v) + C - 1$
6. 从区间[0,1]中均匀采样出一个值 x
7. if $x < \frac{e^{\epsilon/2}}{e^{\epsilon/2}+1}$ then
8. $v \leftarrow$从区间 [l(v),r(v)]均匀采样出v^*
9. else
10. $v \leftarrow$从区间$[-C, l(v)) \cup (r(v), C]$均匀采样出$v^*$
11. $\theta^* = \theta \cdot t$
12. Return θ^*

4 PD-LDPFL 的安全性分析与效用分析

本章首先在理论上分析 PD-LDPFL 的隐私保证以及改进后的 PM 算法的误差。随后，通过理论分析和实验验证，证明即使采用较高的隐私预算，PD-LDPFL 也能够天然地抵御某些隐私推理攻击。最后，通过实验得出了个性化模型的最优设置，并分析了这些设置减少 LDP 引入异构性的原因。

4.1 LDP 保证和方差分析

在 PD-LDPFL 中，每个客户端都拥有一个由个性化模型和基础模型组合而成的拓展模型。个性化模型被保存在本地，因此无需采取额外的隐私保护措施。基础模型则使用 LDP 机制中的 PM 算法对其参数进行扰动。由于 PM 算法满足 LDP 定义的证明可以在 2.2 节引用的参考文献中找到，因此不再赘述。接下来，我们将从 PM 算法的方差入手，证明参数放大会减小最终的 LDP 误差。

定理 1 给定一个模型 θ，经过参数放大后，PM 机制引入的方差可以减少到 $\frac{v^2}{e^{\epsilon/2}-1} + \frac{m^2 \cdot (e^{\epsilon/2}+3)}{3(e^{\epsilon/2}-1)^2}$（$m$ 是模型 θ 中参数的最大值，并且 $0 < m \ll 1$）。

证明：客户端的模型参数 θ 经过 PM 算法扰动后，θ 中的任一值 v 可以表示为 $v + Noise_{PM}(v)$。其中 $Noise_{PM}(v)$ 是一个随机变量，表示 PM 算法引入的误差，如式(4)所示：

$$Var(Noise_{PM}(v)) = \frac{v^2}{e^{\epsilon/2}-1} + \frac{e^{\epsilon/2}+3}{3(e^{\epsilon/2}-1)^2} \tag{4}$$

而经过参数放大和 PM 算法扰动后，v 可以表示为 $\frac{v}{m} + Noise_{PM}$，且：

$$Var\left(Noise_{PM}\left(\frac{v}{m}\right)\right) = \frac{\frac{v^2}{m^2}}{e^{\epsilon/2}-1} + \frac{e^{\epsilon/2}+3}{3(e^{\epsilon/2}-1)^2} \tag{5}$$

由于放大后得到的不是原始值的无偏估计，因此还需要对其放缩，即最终 v 被表示为 $v + m \cdot Noise_{PM}\left(\frac{v}{m}\right)$，此时：

$$Var\left(m \cdot Noise_{PM}\left(\frac{v}{m}\right)\right) = \frac{v^2}{e^{\epsilon/2}-1} + \frac{m^2 \cdot (e^{\epsilon/2}+3)}{3(e^{\epsilon/2}-1)^2} \tag{6}$$

4.2 PD-LDPFL 天然抵抗隐私推理攻击的实验分析

隐私推理攻击通常需要联邦学习客户端的模型参数或模型更新量（梯度）。由于客户端模型经训练集训练后得到的信息是一定的，而 PD-LDPFL 采用了参数解耦的结构设置，将个性化模型保留在本地，只上传基础模型，由于基础模型包含的训练集信息会相对减少，因此 PD-LDPFL 可以在一定程度上抵御某些隐私推理攻击。接下来将用实验验证这个结论。在实验中，使用狄利克雷系数为 5 的非独立同分布的 MNIST 和 CIFAR-10 数据集，客户端的数量设置为 10，训练轮次设置为 50，学习率设置为 0.1。我们仅改变了 LDP 噪声的大小，攻击参数采用了 DLG 本身的设置。简单而言，首先随机选择一个客户端获得其上传的参数信息，然后根据上一轮模型参数与当前轮次模型参数的差值反推出梯度，最后使用与客户端图片像素相同的虚拟图片来训练当前轮次的基础模型并不断添加扰动，直到虚拟图片训练得到的梯度与客户端图片的梯度距离小于某个差值或不再减少。

图 3 和图 4 给出了在 MNIST 和 CIFAR-10 数据集上使用 DLG 方法攻击 Fedavg 和 PD-LDPFL 中间传输梯度的结果。可以观察到，当添加的噪声较弱时，DLG 能够用较少的攻击迭代轮次精确地还原 Fedavg 中参与客户端的隐私图像，而当添加的噪声增加时，DLG 对 Fedavg 的攻击效果就会被削减。相反地，当采用参数解耦的结构设置，即 PD-LDPFL 时，即使添加较少的噪声，DLG 也无法恢复出隐私图像。

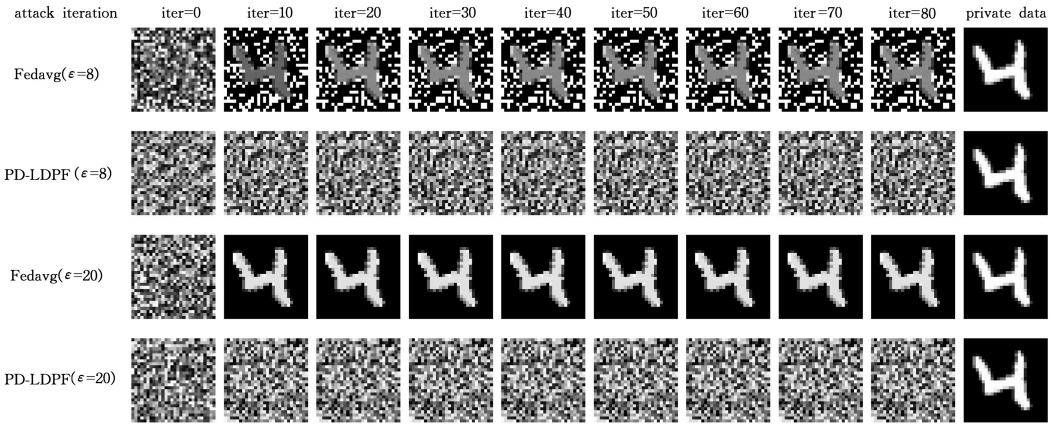


图 3 DLG 攻击在 MNIST 数据集上的实验结果

Fig. 3 Experimental results of DLG attack on MNIST dataset

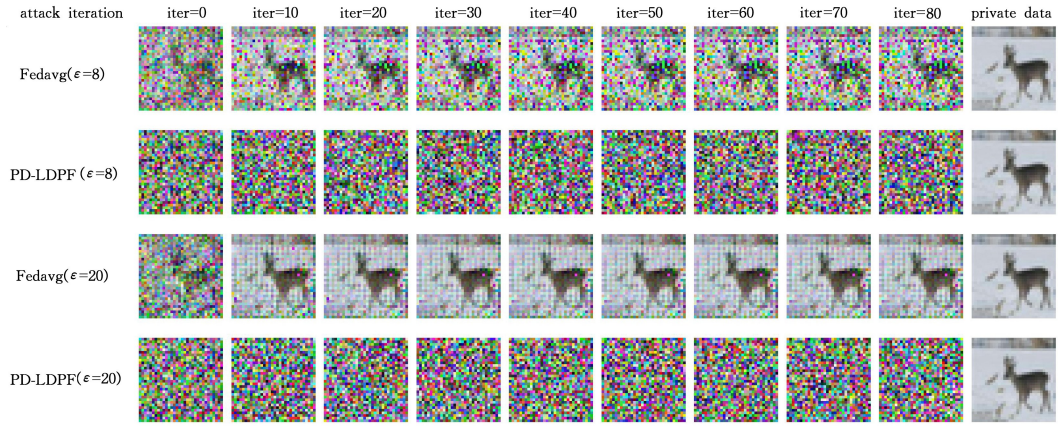


图 4 DLG 攻击在 CIFAR-10 数据集上的实验结果

Fig. 4 Experimental results of DLG attack on CIFAR-10 dataset

4.3 个性化模型设置的实验分析

在个性化输入和输出模型的设置中，包括两个关键部分：变换函数和初始化参数。变换函数决定了个性化模型是否能够克服数据间的异构性并最大化保留原始数据的信息。初始化参数的设置影响着模型的收敛速度。在我们的实验设置中，使用了狄利克雷系数为 0.05 的非独立同分布 MNIST 数据集，客户端数量设置为 10，隐私预算设为 8，训练轮次设置为 50，学习率设置为 0.1。

4.3.1 个性化模型的变换函数设置

理论上，变换函数的空间是无限的，我们为个性化输入模型和个性化输出模型选择了一些常用的变换函数，例如线性变换 $\boldsymbol{\alpha x}+\boldsymbol{b}$，非线性变换 $\log(1+\boldsymbol{\alpha x}+\boldsymbol{b})$，$\tanh(\boldsymbol{\alpha x}+\boldsymbol{b})$，$\mathrm{sigmoid}(\boldsymbol{\alpha x}+\boldsymbol{b})$等。由于实验的输入 $\boldsymbol{x}$ 是 $n\times n\times d$ 维的张量，输出是 m 维张量。为了适配输入和输出，在个性化输入模型中，$\boldsymbol{\alpha}_{\text{input}}$ 设置为一个 d 维的张量，偏置 $\boldsymbol{b}_{\text{input}}$ 设置为 $n\times n\times d$ 的张量。在个性化输出模型中，$\boldsymbol{\alpha}_{\text{output}}$ 设置为一个 1 维的张量，偏置 $\boldsymbol{b}_{\text{output}}$ 设置为 m 维的张量。参数 $\boldsymbol{\alpha}_{\text{input}}$，$\boldsymbol{\alpha}_{\text{output}}$，$\boldsymbol{b}_{\text{input}}$ 和 $\boldsymbol{b}_{\text{output}}$ 的值都是通过学习得到的。表 1 列出了采用个性化模型设置后，客户端间损失函数的累计方差，即异构性。表 1 中累计方差值为 NAN 的变换组合表示模型无法收敛。

表 1 不同个性化模型设置在非独立同分布数据集上的异构性

Table 1 Heterogeneity of different personalized model configurations on Non-I. I. D. distributed dataset

Heterogeneity	$\boldsymbol{x}$	$\boldsymbol{\alpha x}+\boldsymbol{b}$	$\log(1+\boldsymbol{\alpha x}+\boldsymbol{b})$	$\tanh(\boldsymbol{\alpha x}+\boldsymbol{b})$	$\mathrm{sigmoid}(\boldsymbol{\alpha x}+\boldsymbol{b})$
$\boldsymbol{x}$	0.580	0.416	NAN	0.191	0.126
$\boldsymbol{\alpha x}+\boldsymbol{b}$	0.464	0.396	NAN	0.199	0.098
$\log(1+\boldsymbol{\alpha x}+\boldsymbol{b})$	0.524	0.510	NAN	0.165	0.143
$\tanh(\boldsymbol{\alpha x}+\boldsymbol{b})$	0.549	0.526	NAN	0.184	0.124
$\mathrm{sigmoid}(\boldsymbol{\alpha x}+\boldsymbol{b})$	4.114	4.47	NAN	1.86	0.503

从表 1 可以观察到，对于个性化输入模型，采用线性变换、对数变换和双曲正切变换的模型的异构性均小于没有添加个性化输入模型的模型异构性（即 0.58）。然而，sigmoid 变换似乎引入了大量的额外异构性，这是因为 sigmoid 变换对原始输入特征造成了极大的破坏，导致模型无法学习到样本的特性信息。对于个性化输出模型，线性变换、双曲正切变换和 sigmoid 变换都取得了较好的效果，而对数转换会导致模型无法收敛，这是因为对数变换使得模型的输出趋于一致，从而导致模型无法学习到任何知识。图 5 给出了表 1 中较优的个性化变换组合对模型准确率的影响，其中图标为

x&x 的曲线指不添加个性化模型的准确率。尽管在表 1 中，个性化输出模型采用 sigmoid 变换时异构性最小，但图 5 中的结果显示，当个性化输出模型采用 sigmoid 变换时，模型的准确率反而最低，这说明我们在考虑变换降低异构性的同时也要考虑其是否有益于模型收敛。此外，可以观察到，当个性化输入和输出模型均设置为线性变换时，模型的准确率最高，虽然这种设置在表 1 中的异构性不是最低的。这可能是因为当个性化输入和输出模型设置为线性变换时更好地保留了训练样本中的输入和输出特征，同时有效地减小了 LDP 引入的异构性。

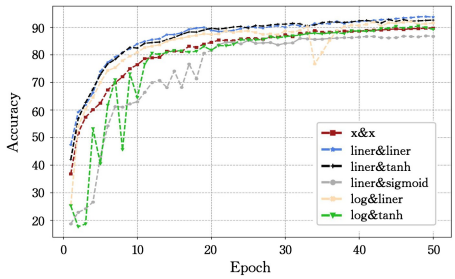


图 5 不同个性化模型设置对 PD-LDPFL 准确性的影响

Fig. 5 Impact of different personalized model configurations on the accuracy of PD-LDPFL

4.3.2 初始化参数分析

在 4.3.1 节中的实验结果表明当个性化输入和输出模型都设置为线性变换时，其效果比其他方案更优。现在将探索线性变换的初始参数，这些初始参数决定了拓展模型的收敛速度。尽管从实验结果来看，在 MNIST 数据集上训练 50 轮后，客户端中无论是输入层还是输出层的线性变换，α 的取值在 1 到 3 之间，b 的取值在 0 附近。我们的直觉认为保持恒等变换对于模型的收敛作用最大，即 α 的初始化值设置为 1，b 的初始化值全部设置为 0。

图 6 给出了在不同参数 α 和 b 初始值下的全局模型的测试精度，为了方便比较，个性化输入和输出模型使用相同的初始化设置。

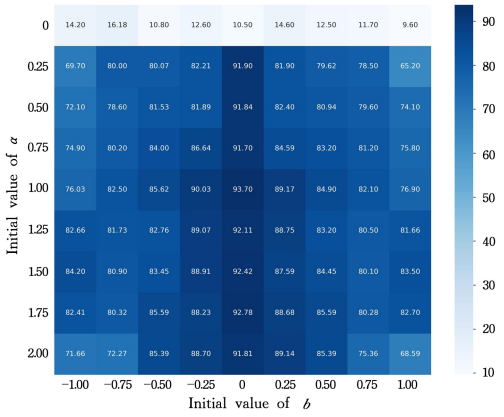


图 6 不同初始化参数设置对 PD-LDPFL 准确性的影响

Fig. 6 Impact of different initialization parameter configurations on the accuracy of PD-LDPFL

具体而言，我们将线性变换的参数 α 的初始值分配在[0，2]范围内，参数 b 的初始值分配在[-1，1]范围内。通过观察图 6，可得如下结论：首先，当 α 取值为 0 时，无论 b 如何取值，拓展模型的准确率始终非常低，这可能是因为当参数 α 初始化为 0 后，拓展模型无法学习到任何有用的信息；其次，当 b 取值接近 0 时，拓展模型的准确率总是在 α 取相同值时最高（排除 α 取值为 0）；最后，当 α 取值为 1，b 取值为 0 时，拓展模型的准确率最高，这可能是因此恒等变换最大程度地保留了原始数据的特征，从而加快了模型的收敛速度。

5 实验

本章对 PD-LDPFL 在多个数据集和深度学习模型中的性能进行了评估，并将其与 Fedavg 和 Privatefl 进行了比较。为了观察 PD-LDPFL 的稳定性和效用，我们进行了一系列实验，涉及改变隐私预算、客户端数量、是否添加个性化模型以及狄利克雷系数等参数。我们观察到，当不使用个性化模型时，PD-LDPFL 的性能会下降至与 Fedavg 相当；而当仅存在个性化输入模型或个性化输出模型时，PD-LDPFL 的性能则会下降至与 Privatefl 相近。

5.1 实验设置

5.1.1 环境配置

实验在 Ubuntu 20.04 环境下进行，使用的编程语言和版本为 Python 3.8。实验的评估指标涵盖 3 个方面，即全局模型的准确率、拓展模型在测试集上的最高准确率以及客户端间损失函数的累计方差。

5.1.2 数据集

本文选用了一系列图像基准数据集，包括 MNIST，Fashion-MNIST 和 CIFAR-10。MNIST 是一个经典的手写数字识别数据集，源自美国国家标准与技术研究所，其中包含了不同人手写的 0 到 9 的数字灰度图片。该数据集由 60000 个训练图片和 10000 个测试图片构成，每个样本的像素为 28×28。Fashion-MNIST 数据集则是来自德国 Zalando 的服装图像数据集，包含 60000 个样本的训练集和 10000 个样本的测试集，每个样本都是与 10 类标签相关的 28×28 像素的灰度图像。另外，CIFAR-10 是一个用于识别普适物体的小型数据集，共含有 10 个类别的 RGB 彩色图像，图像的尺寸为 32×32，数据集中包含 50000 张训练图片和 10000 张测试图片。

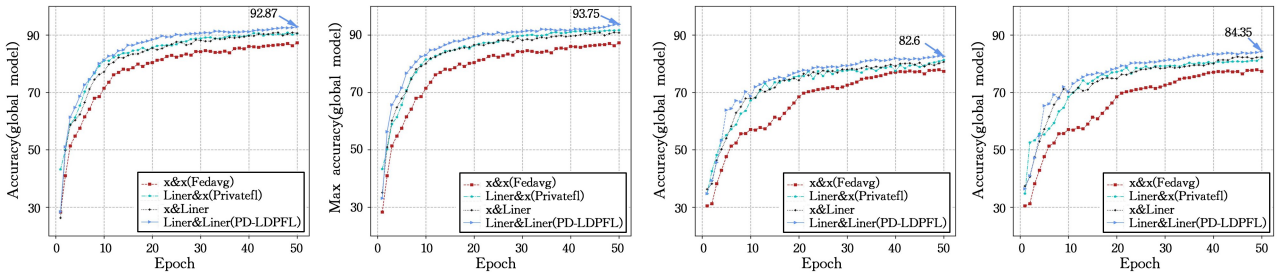
5.1.3 模型结构

本文的基础模型采用了自定义的全连接神经网络模型以及 ResNet18 模型。全连接神经网络模型的结构包括 1 个输入层、2 个隐藏层和 1 个输出层。ResNet18 是一种经典的深度卷积神经网络模型，由微软亚洲研究院提出，其主要特点在于引入了残差块的概念来解决深度卷积神经网络中的梯度消失和梯度爆炸问题。在 ResNet18 中，其网络包含了 18 个卷积层。

5.2 个性化模型设置的影响

个性化模型的设置是 PD-LDPFL 的核心思想。与以往的工作不同，PD-LDPFL 首次采用了个性化输入模型、基础模

型、个性化输出模型相结合的设置。图 7 给出了在 MNIST 和 Fashion-MNIST 数据集上使用全连接神经网络基础模型进行实验得到的全局模型准确率和拓展模型的最高准确率。在实验中，我们将客户端数量设为 10，狄利克雷系数设为 0.05，学习率设为 0.05，隐私预算设为 8。图中的红色虚线代表未添加任何个性化模型(Fedavg)，而青色虚线代表添加了个性化输入转换(Privatefl)。首先，观察图 7 中 4 张子图前 10 轮的曲线差异，可以发现 PD-LDPFL 的收敛速度相对于其他方法更为迅速，而 Privatefl 的收敛速度和 Fedavg 差不多。这表明我们在引言部分的判断是正确的，即 Privatefl 的个性化转换参数在开始时学习较慢。其次，在 50 轮训练中，PD-LDPFL 的准确率均显著高于 Fedavg 和 Privatefl。这说明本文方案可以更好地克服异构性并减少 LDP 造成的精度损失。此外，在添加个性化模型后，每个客户端都拥有独特的拓展模型，如图 7(b)和图 7(d)所示，拓展模型在测试集上的最高准确率显著高于全局模型。这意味着使用 PD-LDPFL 的客户端可以获得个性化的解决方案，其中大部分客户端的拓展模型的效果优于全局模型，并且即使在最差情况下，小部分客户端也能获得与全局模型相当的效果。另一个值得注意的发现是，当仅添加个性化输入模型和仅添加个性化输出模型时，全局模型的准确率曲线非常接近，这可能是因为它们在平衡异构性方面具有相似的效果。



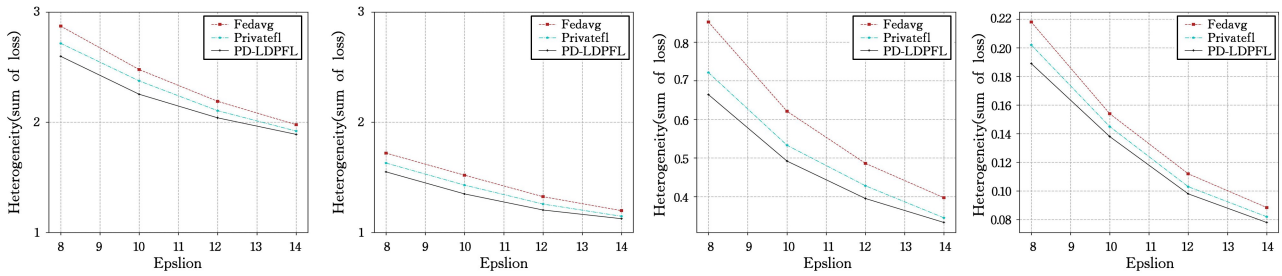
(a)MNIST dataset (b)MNIST dataset (c)FMNIST dataset (d)FMNIST dataset

图 7 不同联邦学习方法在非独立同分布的 MNIST 和 Fashion MNIST 上的准确性实验(电子版为彩图)

Fig. 7 Experiments on the accuracy of different FL methods on Non-I. I. D. MNIST and Fashion MNIST datasets

5.3 非独立同分布和隐私预算的影响

客户端间数据的异构性和噪声添加的大小严重影响着联邦学习的收敛效率。为了深入分析异构性和噪声对 PD-LDPFL 的影响，我们进行了一系列实验，比较了 4 种不同数据分布下 PD-LDPFL，FedAvg 和 PrivateFL 的表现。针对 CIFAR-10 数据集，选择了不同的狄利克雷系数，分别为 0.5，1，5 和 100，以评估在多种隐私预算下客户端间损失函数的累计方差。在实验设置中，固定客户端数量为 10，学习率为 0.05，全局训练轮次为 50，并将全局模型初始化为 ResNet18，同时将所有个性化模型设定为线性变换。如图 8 所示，首先可以发现当隐私预算增加时客户端间的异构性减少，且 3 种方法的异构性相对差也减小。这是因为隐私预算控制着添加的噪声大小，隐私预算越大，添加的噪声越小。其次，无论是何种数据分布，PD-LDPFL 均取得了最佳效果，有效地降低了客户端之间的异构性。最后，结合 4 种数据分布下客户端损失函数值累计方差的相对差值，即图 8(a)中 PD-LDPFL 和 Fedavg 在隐私预算为 8 时差值约为 0.2，图 8(d)中差值约为 0.03，说明当数据集的非独立同分布特性越强时，PD-LDPFL 的表现越为出色。



(a)β=0.5 (b)β=1 (c)β=5 (d)β=100

图 8 数据异构性和隐私预算在非独立同分布的 CIFAR-10 数据集上对不同联邦学习方法的影响

Fig. 8 Impact of data heterogeneity and privacy budget on different FL methods on Non-I. I. D. CIFAR-10 dataset

5.4 客户端数量的影响

本文在 MNIST 和 FMNIST 数据集上对 PD-LDPFL 进行了在不同客户端数量下的性能影响研究。在实验中，我们将隐私预算设定为 8，学习率设为 0.05，全局训练轮次设为 50。客户数量范围从 5 到 50，并且将狄利克雷系数设为 1，全局模型初始化为全连接神经网络模型，并将所有个性化模型设置为线性变换。

表 2 列出了随着客户端数量变化，PD-LDPFL，Fedavg 和 PrivateFL 在测试集上的全局模型的准确率。我们注意到，无论客户端数量如何变化，PD-LDPFL 的性能始终优于 Fedavg 和 PrivateFL，特别是随着客户端数量增加，3 种方法之间的准确率差距更为显著。此外，随着客户端数量的增加，3 种

方法在这2个数据集上的准确率都呈下降趋势。我们推测这可能是由于在数据集数量固定的情况下，尽管更多客户端有助于平衡本地化差分隐私的误差，但同时也引入了过多的异构性。

表2 客户数量在MNIST和FMNIST数据集上对PD-LDPFL的影响

Table 2 Impact of the number of clients on PD-LDPFL on MNIST and FMNIST datasets

(%)

Number of Clients	MNIST			FMNIST		
	Fedavg	Privatefl	PD-LDPFL	Fedavg	Privatefl	PD-LDPFL
5	92.77	94.80	96.57	80.16	81.26	82.27
10	92.50	95.11	96.40	75.05	75.84	76.32
25	90.29	93.7	95.15	76.68	78.21	79.85
50	89.69	92.97	94.87	74.50	77.31	78.53

结束语 保护参与方的隐私是联邦学习的重要目标，而现有的差分隐私联邦学习方案大多存在精度损失严重、实用性差等问题。因此，本文提出了差分隐私保护下基于参数解耦的联邦学习方案。该方法可以显著减少LDP造成的精度损失，并且相对于传统的差分隐私联邦学习方案具有更高的安全性。未来，我们将继续拓展该方法。我们推测，在本地保留更复杂的个性化模型可能会更大程度提升精度。然而，当本地客户端保留过多信息量时，基础模型的信息量就会大幅减少，必然会减缓模型的收敛速度，这之间应该存在一个权衡。

参 考 文 献

[1] TANG P, XU H M, MA C. ProtoTransfer: Cross-Modal Prototype Transfer for Point Cloud Segmentation[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 3337-3347.

[2] ZHAN F, YU Y, WU R, et al. Marginal contrastive correspondence for guided image generation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 10663-10672.

[3] LEE P, BUBECK S, PETRO J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine[J]. New England Journal of Medicine, 2023, 388(13): 1233-1239.

[4] 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.

[5] PHONG L T, AONO Y, HATASHI T, et al. Privacy-preserving deep learning: Revisited and enhanced[C]// Applications and Techniques in Information Security: 8th International Conference. 2017: 100-110.

[6] WEI W, LIU L, LOPER M, et al. A framework for evaluating gradient leakage attacks in federated learning[J]. ESORICS 2020: 25th European Symposium on Research in Computer Security, 2020, 12308: 545-566.

[7] AONO Y, HAYASHI T, WANG L, et al. Privacy-preserving deep learning via additively homomorphic encryption[J]. IEEE Transactions on Information Forensics and Security, 2017, 13(5): 1333-1345.

[8] YIN H, MALLYA A, VAHDAT A, et al. See through gradients: Image batch recovery via gradinversion[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 16337-16346.

[9] GEIPING J, BAUERMEISTER H, DRÖGE H, et al. Inverting gradients-how easy is it to break privacy in federated learning?[J]. Advances in Neural Information Processing Systems, 2020, 33: 16937-16947.

[10] MOTHUKURI V, PARIZI R M, POURIYEH S, et al. A survey on security and privacy of federated learning[J]. Future Generation Computer Systems, 2021, 115: 619-640.

[11] DWORK C, MCSHERRY F, NISSIM K, et al. Calibrating noise to sensitivity in private data analysis[C]// Thirty-Ninth ACM Symposium on Theory of Computing ACM. 2007: 75-84.

[12] DWORK C, ROTH A. The algorithmic foundations of differential privacy[J]. Foundations and Trends in Theoretical Computer Science, 2014, 9(3/4): 211-407.

[13] KASIVISWANATHAN S P, LEE H K, NISSIM K, et al. What can we learn privately? [J]. SIAM Journal on Computing, 2011, 40(3): 793-826.

[14] KIM M, JAIN A K, LIU X. Adaface: Quality adaptive margin for face recognition[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 18750-18759.

[15] XU J, HE X, LI H. Deep learning for matching in search and recommendation[C]// The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. 2018: 1365-1368.

[16] WANG T, HU X, LIU Z, et al. Sparse2Dense: Learning to densify 3d features for 3d object detection[J]. Advances in Neural Information Processing Systems, 2022, 35: 38533-38545.

[17] YE M, FANG X, DU B, et al. Heterogeneous federated learning: State-of-the-art and research challenges[J]. ACM Computing Surveys, 2023, 56(3): 1-44.

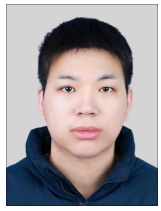
[18] FU J, CHEN Z, HAN X. Adap DP-FL: Differentially Private Federated Learning with Adaptive Noise[C]// 2022 IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). IEEE, 2022: 656-663.

[19] HUANG X, DING Y, JIANG Z L, et al. DP-FL: a novel differentially private federated learning framework for the unbalanced data[J]. World Wide Web, 2020, 23: 2529-2545.

[20] YANG Y, HUI B, YUAN H, et al. PrivateFL: Accurate, Differentially Private Federated Learning via Personalized Data Transformation[C]// 32nd USENIX Security Symposium (USENIX Security 23). 2023: 1595-1612.

[21] LI Q, WEN Z, WU Z, et al. A survey on federated learning systems: Vision, hype and reality for data privacy and protection

[J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 35(4): 3347-3366.
[22] 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.
[23] 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.
[24] SAJADMANESH S, SHAMSABADI A S, BELLET A, et al. Gap: Differentially private graph neural networks with aggregation perturbation[C]// 32nd USENIX Security Symposium. 2023.
[25] SUN L, QIAN J, CHEN X. LDP-FL: Practical Private Aggregation in Federated Learning with Local Differential Privacy[C]// Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence Organization, 2021.
[26] WANG N, XIAO X, YANG Y, et al. Collecting and analyzing multidimensional data with local differential privacy[C]// 2019 IEEE 35th International Conference on Data Engineering (ICDE). IEEE, 2019: 638-649.
[27] DUCHI J C, JORDAN M I, WAINWRIGHT M J. Minimax optimal procedures for locally private estimation[J]. Journal of the American Statistical Association, 2018, 113(521): 182-201.
[28] BU Z, DONG J, LONG Q, et al. Deep Learning with Gaussian Differential Privacy[J]. Harvard Data Science Review, 2020.
[29] 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. 2016: 308-318.
[30] WANG N, XIAO X, YANG Y, et al. Collecting and analyzing multidimensional data with local differential privacy[C]// 2019 IEEE 35th International Conference on Data Engineering (ICDE). IEEE, 2019: 638-649.
[31] ZHU L, LIU Z, HAN S. Deep leakage from gradients[C]// Advances in Neural Information Processing Systems. 2019: 14747-14756.
[32] KARIMIREDDY S P, KALE S, MOHRI M, et al. Scaffold: Stochastic controlled averaging for federated learning[C]// International Conference on Machine Learning. PMLR, 2020: 5132-5143.
[33] TAN Y, LONG G, LIU L, et al. Fedproto: Federated prototype learning across heterogeneous clients[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2022: 8432-8440.
[34] LI Q, DIAO Y, CHEN Q, et al. Federated learning on non-iid data silos: An experimental study[C]// 2022 IEEE 38th International Conference on Data Engineering (ICDE). IEEE, 2022: 965-978.
[35] SATTLER F, MÜLLER K R, SAMEK W. Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 32(8): 3710-3722.
[36] WU Q, HE K, CHEN X. Personalized federated learning for intelligent IoT applications: A cloud-edge based framework[J]. IEEE Open Journal of the Computer Society, 2020, 1: 35-44.
[37] HÖNIG R, ZHAO Y, MULLINS R. DAdaQuant: Doubly-adaptive quantization for communication-efficient Federated Learning[C]// International Conference on Machine Learning. PMLR, 2022: 8852-8866.
[38] WANG Y, LIN L, CHEN J. Communication-efficient adaptive federated learning[C]// International Conference on Machine Learning. PMLR, 2022: 22802-22838.
[39] LI T, SAHU A K, TALWALKAR A, et al. Federated learning: Challenges, methods, and future directions[J]. IEEE Signal Processing Magazine, 2020, 37(3): 50-60.
[40] NIELSEN M A. Neural networks and deep learning [M]. San Francisco, CA, USA: Determination Press, 2015.



WANG Zihang, born in 1999, postgraduate, is a member of CCF (No. R6802G). His main research interests include differential privacy and federated learning.

YANG Min, born in 1975, Ph. D, associate professor, master supervisor, is a member of CCF (No. 51131M). Her main research interests include information security and applied cryptography.

(责任编辑:喻藜)