

动态聚合权重的隐私保护联邦学习框架

应作斌¹, 方一晨¹, 张怡文²

(1. 澳门城市大学, 中国 澳门 999078; 2. 安徽新华学院, 安徽 合肥 230000)

摘要: 在非可信中心服务器下的隐私保护联邦学习框架中, 存在以下两个问题。① 在中心服务器上聚合分布式学习模型时使用固定的权重, 通常是每个参与方的数据集大小。然而, 不同参与方具有非独立同分布的数据, 设置固定聚合权重会使全局模型的效用无法达到最优。② 现有框架建立在中心服务器是诚实的假定下, 没有考虑中央服务器不可信导致的参与方的数据隐私泄露问题。为了解决上述问题, 基于比较流行的 DP-FedAvg 算法, 提出了一种非可信中心服务器下的动态聚合权重的隐私保护联邦学习 DP-DFL 框架, 其设定了一种动态的模型聚合权重, 该方法从不同参与方的数据中直接学习联邦学习中的模型聚合权重, 从而适用于非独立同分布的数据环境。此外, 在本地模型隐私保护阶段注入噪声进行模型参数的隐私保护, 满足不可信中心服务器的设定, 从而降低本地参与方模型参数上传中的隐私泄露风险。在数据集 CIFAR-10 上的实验证明, DP-DFL 框架不仅提供本地隐私保证, 同时可以实现更高的准确率, 相较 DP-FedAvg 算法模型的平均准确率提高了 2.09%。

关键词: 联邦学习; 差分隐私; 动态聚合权重; 非独立同分布数据

中图分类号: TP393

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2022069

Privacy-preserving federated learning framework with dynamic weight aggregation

YING Zuobin¹, FANG Yichen¹, ZHANG Yiwen²

1. City University of Macau, Macau 999078, China

2. Anhui Xinhua University, Hefei 230000, China

Abstract: There are two problems with the privacy-preserving federal learning framework under an unreliable central server. ① A fixed weight, typically the size of each participant's dataset, is used when aggregating distributed learning models on the central server. However, different participants have non-independent and homogeneously distributed data, then setting fixed aggregation weights would prevent the global model from achieving optimal utility. ② Existing frameworks are built on the assumption that the central server is honest, and

收稿日期: 2022-04-09; 修回日期: 2022-09-06

通信作者: 方一晨, fyc980601@163.com

基金项目: 澳门科学技术发展基金 (0038/2022/A)

Foundation Item: General R&D Subsidy Program Fund Macau (0038/2022/A)

引用格式: 应作斌, 方一晨, 张怡文. 动态聚合权重的隐私保护联邦学习框架[J]. 网络与信息安全学报, 2022, 8(5): 56-65.

Citation Format: YING Z B, FANG Y C, ZHANG Y W. Privacy-preserving federated learning framework with dynamic weight aggregation [J]. Chinese Journal of Network and Information Security, 2022, 8(5): 56-65.

do not consider the problem of data privacy leakage of participants due to the untrustworthiness of the central server. To address the above issues, based on the popular DP-FedAvg algorithm, a privacy-preserving federated learning DP-DFL algorithm for dynamic weight aggregation under a non-trusted central server was proposed which set a dynamic model aggregation weight. The proposed algorithm learned the model aggregation weight in federated learning directly from the data of different participants, and thus it is applicable to non-independent homogeneously distributed data environment. In addition, the privacy of model parameters was protected using noise in the local model privacy protection phase, which satisfied the untrustworthy central server setting and thus reduced the risk of privacy leakage in the upload of model parameters from local participants. Experiments on dataset CIFAR-10 demonstrate that the DP-DFL algorithm not only provides local privacy guarantees, but also achieves higher accuracy rates with an average accuracy improvement of 2.09% compared to the DP-FedAvg algorithm models.

Keywords: federated learning, differential privacy, dynamic aggregation weight, non-independent and identically distributed data

0 引言

随着大数据的兴起与发展,机器学习在计算机视觉、医疗、金融、电信、无人驾驶等领域广泛应用^[1]。传统的机器学习是由一个服务商从各个参与用户处收集数据进行统一的模型训练,因此会直接将原始数据泄露给第三方服务商。而这类数据中可能包含敏感类信息,如医疗健康信息(个人生理特征、既往病史、基因序列等)、文本邮件、交易记录、资产数据、行踪轨迹等^[2-3]。

在满足数据融合需求的条件下,解决数据孤岛难题,同时保障包含敏感信息的数据共享中的隐私安全,联邦学习技术应运而生。联邦学习最早在2016年由Google公司提出,是一种分布式的机器学习框架,其核心机制是多个数据持有方(个人、企业、机构)在参与协同训练时,仅在中间过程中交换模型参数,无须直接共享原始数据,从而避免原始敏感信息的泄露^[4-6]。联邦学习在一定程度上有效保护了原始数据直接泄露给第三方^[7-9],但随着对联邦学习的深入研究,其自身框架存在的安全问题被挖掘出来^[10-12]。首先,联邦学习存在一个中心服务器,中心服务器接收来自所有本地参与方客户端的模型参数,Nicolas等^[13]发现,一个不可信的中心服务器可以通过共享的参数(如梯度)还原原始的数据集。其次,参与联邦学习的不可靠参与方客户端可以轻松并且合法得到中间参数从而推断出其他参与方的模型参数,进而通过模型参数获取用户数据^[14]。参与方客户端也可以通过构建恶意的数据集诱导其

他参与方暴露更多含有敏感信息的原始数据。此外,其他的第三方攻击者可以通过发布的模型进行模型重构演练从而还原原始数据集收集敏感信息^[15]。

对于联邦学习算法的隐私保护问题,需要一个严格的、有原则的框架来加强数据隐私。以差分隐私为代表的扰动方法作为成熟的隐私保护技术已广泛运用于联邦学习^[16-23]。已有在联邦学习的框架上引入差分隐私从而保护用户数据的研究,McMahan等^[24]提出了DP-FedAvg(differentially private-federated averaging)算法。DP-FedAvg在联邦学习模型参数共享算法FedAvg^[4]的基础上,引入了差分隐私,是第一个通过应用高斯机制来保证联邦学习中用户级差分隐私的算法。DP-FedAvg算法主要有两个阶段:①参与方使用本地数据集训练模型;②中心服务器收集并聚合模型,通过加权平均获得全局模型并注入高斯噪声。然而,DP-FedAvg算法存在以下两个问题。

1) 在中心服务器中全局模型的聚合权重通常由用户端的数据样本量决定,这种设计假定数据在不同客户端之间是同分布的,但在现实世界中,无法知道不同参与方的数据分布类型,即不同参与方之间很可能是非独立同分布的数据(non-iid, non-independent and identically distributed data)。采用此种聚合方法导致模型的效用和适用能力不足。

2) DP-FedAvg算法建立在中心服务器是诚实的假定下,然而中心可以接收到来自所有本地参与方的模型参数信息,一旦中心服务器不可信,

它能利用本地参与方的模型参数倒推数据的分布特征,甚至倒推出具体的训练集数据,导致本地参与方的数据隐私泄露,甚至出于利益相关将本地参与方的数据售卖给其他第三方,进一步导致本地参与方的数据隐私大规模泄露。

针对以上问题,本文提出了一种基于非可信中心服务器设定下的动态聚合权重的隐私保护联邦学习(DP-DFL, differentially private- dynamic federated learning)框架,主要贡献总结如下。

1) 构建了一种动态的模型聚合权重的方法,该方法可以从数据中直接学习联邦学习中的模型聚合权重,进一步实现更好的性能和不同数据分布类型的通用。

2) 假定中心服务器不可信,设计了一种新的协议,参与方客户端在本地模型隐私保护阶段使用噪声进行模型参数的隐私保护,从而降低本地参与方模型参数上传中的隐私泄露风险。

3) 验证了所提算法的效率,在数据集 CIFAR-10 上采用两种卷积神经网络模型进行实验,结果显示,DP-DFL 在有效保证本地隐私的条件下,平均准确率相比 DP-FedAvg 算法提高了 2.09%。

1 相关工作

越来越多的工作研究联邦学习背景下的隐私保护,将 DP 约束引入联邦学习模型共享是其中一种主要的研究。Geyer^[25]等提出了一种算法,中心服务器将高斯机制注入每一轮迭代后的权重矩阵中,将单个客户的贡献隐藏在聚合中。

McMahan^[24]等也提出了类似的想法,DP-FedAvg 算法和 DP-FedSGD 算法扩展了 FedAvg 和 FedSGD。FedAvg^[4]算法和 FedSGD^[4]算法是联邦学习中的经典算法,中心服务器聚合每一轮迭代本地参与方客户端的最优梯度更新,FedAvg 算法与 FedSGD 算法相比,前者在本地参与方客户端进行多轮迭代。DP-FedAvg 算法和 DP-FedSGD 算法则是在每一轮的模型更新中加入扰动。以上算法在共享模型参数时通常假定不同参与方的数据具有相同分布,在异质环境下,DP-FedAvg 算法和 DP-FedSGD 算法存在不稳定性收敛的问题。

除此之外,非独立同分布数据上的联邦学习

问题引起了人们的关注。在非 DP 环境下,Wang^[26]提出了一种归一化的平均方法,在保持异质数据用户端快速收敛的同时,消除了目标的不一致性。Zheng^[27]等在共享参数平均化阶段引入了基于注意力的加权平均法,以针对异质数据。在 DP 环境下,Hu^[28]等针对数据异质性问题,提出了一种个性化的 FL-DP 方法,但此方法只针对线性模型。现有工作通常是在非 DP 环境下,考虑模型的统计数据或基本分布进行聚合,对客户数据进行基于梯度的优化,直接学习聚集权重的研究较少。联邦学习中 DP 约束条件下数据异质环境问题的研究尚且不足,尤其是针对神经网络模型及数据类型为图片的原始数据。为了解决数据异质性这一问题,在 DP 约束条件下的联邦学习中,本文引入了一种为联邦学习寻找最佳聚合权重的自动化方法。通过使用 Dirichlet 分布^[29-30]对权重进行建模,将聚合权重由原来的依据客户端数据集大小优化为依据不同参与方的数据分布和模型的训练进度。

上述研究通常假定中心服务器是诚实的,然而实际中,中心服务器可能是“诚实但好奇”或是“恶意”的。Zhao^[31]等考虑了 DP-FedSGD 算法在“诚实但好奇”的中心服务器条件下的联邦学习隐私保护。Li^[32]等认为服务器是“诚实而好奇”的,但没有确保用户的端到端隐私。本文基于 DP-FedAvg 算法,并且假设中心服务器状态为非可信,设定中心服务器处于两种状态中的“诚实但好奇”这一状态,本地参与方客户端在将参数发送给中心服务器前需要进行扰动保护。相关工作比较如表 1 所示。

表 1 相关工作比较
Table 1 Comparison of related work

方法来源	隐私保护	针对 non-iid 全局模型权重		设定中心服务器状态
		数据	聚合方式	
FedAvg ^[4]	否	否	数据集大小	可信
DP-FedSGD ^[24]	是	否	数据集大小	可信
DP-FedAvg ^[24]	是	否	数据集大小	可信
Wang20 ^[26]	否	是	数据分布	可信
Zheng21 ^[27]	否	是	数据分布	可信
FL-DP ^[28]	是	是	数据集大小	可信
Zhao20 ^[31]	是	否	数据集大小	非可信
本文	是	是	动态权重聚合	非可信

2 基本定义

2.1 联邦学习

假设联邦学习框架中有 p 个本地参与方，第 i 个参与方持有一组本地用户数据集 S^i ，参与方 i 在本地训练模型，并将模型的参数传输给中心服务器，中心服务器聚合后更新全局模型，并将相应的模型参数返回给一部分参与方，进行新一轮的模型更新。联邦学习通过最小化中心服务器的损失函数，直至达到指定的迭代次数或模型准确率。中心服务器的损失函数为

$$f(w, S) = \sum_{i=1}^p \frac{n_i}{n} f_i(w, S^i)$$

其中， $f_i(w, S^i)$ 表示选取第 i 个本地参与方选取数据集 S^i 样本生成模型 w 的损失函数， n_i 表示每个本地参与方的数据量， $n = \sum_{i=1}^p n_i$ 表示所有本地参与方的数据量总和。

2.2 差分隐私

定义 1 严格差分隐私

对于一个随机算法 $\mathcal{K}$ ，假设其输出任意的一个子集为 S ，如果随机算法 $\mathcal{K}$ 在只有一条数据不同的相邻数据集 D 和 D' 上的输出满足以下条件，则称算法满足 ε 的差分隐私。

$$\Pr[\mathcal{K}(D) = S] \leq \exp(\varepsilon) \Pr[\mathcal{K}(D') = S]$$

其中， ε 非负，称作隐私预算，是随机算法 $\mathcal{K}$ 的隐私保护程度，隐私预算 ε 的值越小，表示隐私保护程度越高。

传统的 DP 基于最严格的假设，即攻击者拥有除了某一条信息以外的所有其他信息，但在现实中这类攻击者是十分少见的。严格差分隐私在隐私性的保护方面十分严格，影响了数据在使用过程中的可用性，在实际应用中主要使用含有松弛机制的差分隐私定义。带有松弛机制的差分隐私与传统差分隐私相比，引入了一个松弛项。

定义 2 松弛差分隐私

对于一个随机算法 $\mathcal{K}$ ，假设其输出任意的一个子集为 S ，如果随机算法 $\mathcal{K}$ 在只有一条数据不同的相邻数据集 D 和 D' 上的输出满足以下条件，则称算法满足 (ε, δ) 的差分隐私。

$$\Pr[\mathcal{K}(D) = S] \leq \exp(\varepsilon) \Pr[\mathcal{K}(D') = S] + \delta$$

其中， ε 非负，称作隐私预算，是随机算法 $\mathcal{K}$ 的隐私保护程度，隐私预算 ε 的值越小，表示隐私保护程度越高。 δ 为松弛项，表示可以容忍的“失败概率”，通常为 $\frac{1}{n^2}$ （其中 n 表示数据集 D 的大小）或更小。隐私预算 ε 的值越小，表示隐私保护程度越高。

2.3 高斯机制

差分隐私通常通过添加噪声的方式来扰动数据，具体实现机制有高斯机制、指数机制、拉普拉斯机制。其中，拉普拉斯机制仅针对传统的差分隐私，而高斯机制可以满足松弛的本地差分隐私。

定义 3 全局敏感度

对于任意的查询函数 $f: \mathbb{R}^D \rightarrow \mathbb{R}^d$ ， D 为输入的数据集， $\mathbb{R}^d$ 为函数输出的 d 维向量，则函数 f 的全局敏感度满足

$$\Delta_f = \max_{D, D'} \|f(D) - f(D')\|_2$$

其中， D 和 D' 是只有一条数据不同的相邻数据集，敏感度衡量了改变数据集中任意一条记录对查询结果造成的最大范围的变化。噪声的大小与全局敏感度有关，全局敏感度越小，实现差分隐私需要注入的噪声越小。

定义 4 高斯噪声机制

对于任意的查询函数 $f: \mathbb{R}^D \rightarrow \mathbb{R}^d$ ， D 为输入的数据集， $\mathbb{R}^d$ 为函数输出的 d 维向量，高斯噪声机制是给 f 的真实输出添加分布为 $N(0, \sigma^2)$ 的噪声，高斯机制的输出 $A(D)$ 满足 (ε, δ) 的差分隐私。

$$A(D) = f(D) + N(0, \sigma^2)$$

其中 $\varepsilon \in (0, 1)$ ，高斯噪声是均值为 0，协方差为 σ^2 的高斯分布， $c^2 = 2 \ln \left(\frac{1.25}{\delta} \right)$ ， $\sigma \geq c \frac{\Delta_f}{\varepsilon}$ 。

2.4 差分隐私随机梯度下降法

定义 $\nabla l(w_k; d)$ 为参数 w_k 在训练数据集 d 上的损失函数，训练的目的在于找到产生尽可能小损失时的参数 w_k 。在实际计算中，通常采用随机梯度下降法（SGD, stochastic gradient descent）进行求解。在 SGD 基础上引入 DP（即差分隐私）

随机梯度下降法 (DP-SGD)。具体来说, 设定用户数据库 S^k , 随机选择 data batch, data batch 定义为随机梯度下降中的小批量样本, 使用批量大小 (batch size) 表示小批量的大小, 在 DP-SGD 中, 定义 batchsize 为 B_k , $I \subseteq \{1, 2, \dots, |S^k|\}$ 共计 B_k 个样本量, 定义 B_k 个 batch 中每个数据 $d \in I$ 的损失函数 ∇l 为梯度 g_d , 即

$$g_d = \nabla l(w_k; d)$$

得到梯度后对梯度进行裁剪, 即

$$g_d = g_d / \max\{1, \|g_d\|/C\}$$

其中, C 为梯度规范约束, 裁剪梯度主要是为了

限制单个数据对整体的影响。随后向每个梯度上加入高斯噪声后计算平均梯度, 沿着梯度的反方向满足损失最小化更新参数 w_k , 即

$$w_k = w_k - \frac{\eta}{B_k} \left(\sum_{d \in I} g_d + N(0, C^2 \sigma^2) \right)$$

其中, η 为学习率, σ^2 为注入的噪声规模。

3 方案实现

3.1 方案概述

本文提出了一种基于非可信中心服务器设定下的动态聚合权重的隐私保护联邦学习框架, 流程如图 1 所示。

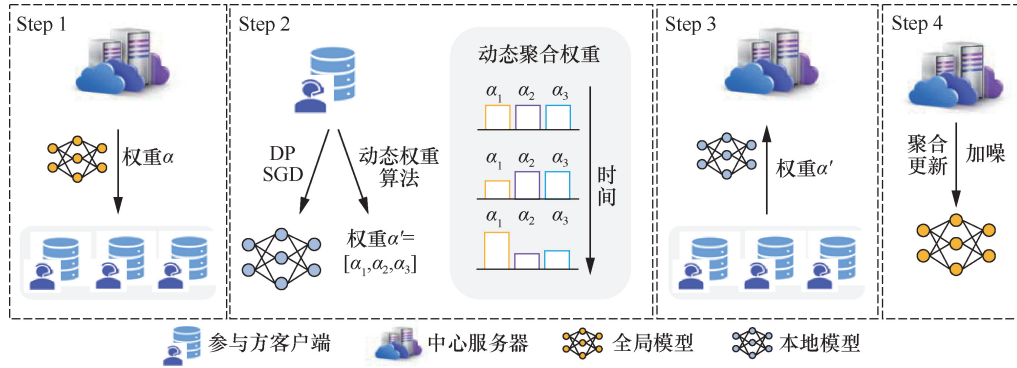


图 1 DP-DFL 流程

Figure 1 The illustration of DP-DFL

DP-DFL 框架的完整流程为: ① 部分本地参与方客户端从中心服务下载全局模型参数和聚合权重参数; ② 本地参与方客户端在本地训练模型和更新权重, 并将 DP 隐私保护, 即采用 DPSGD 方法添加噪声后的模型更新和动态权重算法更新后的聚合权重上传给中心服务器; ③ 中心服务器接受本地参与方客户端的模型更新和聚合权重更新, 生成新的全局模型并进行 DP 隐私保护添加噪声; ④ 重复上述每一轮迭代, 直至达到设定的迭代轮数或期望的模型精确度。

3.2 动态聚合权重算法

DP-FedAvg 算法中, 全局模型的聚合权重通常是固定的, 由参与方数据的样本数据量决定。假设 K 个参与方客户端协同训练一个参数为 w 的全局模型, 迭代多次全局模型满足以下函数。

$$\min \sum_{k=1}^K \alpha_k \mathcal{L}_k(w)$$

其中, $\mathcal{L}_k(w)$ 是 K 个参与方的本地损失函数, 聚合权重设定 $\alpha_k = \frac{n_k}{n}$, $n = \sum_{k=1}^K n_k$ 是参与方的全部样本数据量。此时, 聚合权重是基于数据在参与方客户端间均匀分布的假设。然而, 由于每个参与方客户端的数据分布是未知的, 可能是 non-iid, 这个假设并不适用。

为了解决 DP-FedAvg 算法的局限性, 本文提出了一种从客户的数据中直接学习聚合权重 α 的动态聚合权重算法。将迭代多次全局模型的约束函数优化为

$$\min \sum_{k=1}^K \mathcal{L}_k \left(\sum_{k=1}^K \alpha_k w_k \right)$$

其中, $\sum_{k=1}^K \alpha_k = 1$, $\alpha_k > 0$, w_k 是更新后的本地模型, $w_k = \operatorname{argmin} \mathcal{L}_k(w)$ 。优化后的约束函数可以直接通过梯度下降在数据中以可区分的方式学习聚合

权重。

新建一组新的变量 $\beta = [\beta_1, \beta_2, \dots, \beta_K]$ ， β 与 $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_K]$ 具有相同的向量长度。

定义如下函数进行 α 和 β 之间的转换。

$$\alpha = \gamma(\beta)$$

γ 的实际实例化使用 Dirichlet 分布, Dirichlet 分布中后验概率的密度函数与先验概率的密度函数具有相同的函数形式。具体表达为 $\alpha \sim \text{Dir}(\beta)$ ，其中聚合权重 α 为随机变量， β 为浓度参数的 Dirichlet 分布进行建模。概率密度函数的表达如下。

$$\text{Dir}(\alpha|\beta) = \frac{1}{A(\beta)} \prod_{k=1}^K a_k^{\beta_k-1}$$

其中， $A(\beta) = \frac{\prod_{k=1}^K \Gamma(\beta_k)}{\Gamma(\sum_{k=1}^K \beta_k)}$ ， $\Gamma(z)$ 是伽玛函数，

$$\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x} dx。$$

在训练阶段，随机抽选一个数据样本 x 。在聚合权重学习过程中模型参数 w_k 固定不变，损失函数只取决于 β 和 x ，因此，损失函数 $l\left(\sum_{k=1}^K \alpha_k w_k; x\right) = \mathcal{L}(\beta; x)$ 。利用隐式重参数化^[30]，通过损失函数 $\mathcal{L}(\beta; x)$ 得到浓度参数 β ，从而通过 γ 函数计算得到聚合权重 α 。

算法 1 给出了动态聚合权重算法的描述。

算法 1 动态聚合权重算法

初始化: $\beta^0 = [\beta_1^0, \beta_2^0, \dots, \beta_K^0]$

1) **for** $t = 1, 2, 3, \dots, T$ **do**

2) **for** $k = 1, 2, 3, \dots, K$ **in parallel do**

3) 中心服务器发送 β^{t-1} 给参与方 k

4) 参与方 k 随机挑选一批数据量为 x 的数据

5) 参与方 k 计算损失函数 $\mathcal{L}(\beta^{t-1}; x)$

6) 计算梯度并更新参数 $\beta^{t,k} \leftarrow \beta^{t-1}$

7) 参与方 k 发送 $\beta^{t,k}$ 给中心服务器

8) $\beta^t \leftarrow \frac{1}{K} \sum_{k=1}^K \beta^{t,k}$

9) **end for**

10) **return** β^t

3.3 DP-DFL 框架

表 2 列出了 DP-DFL 框架相关符号和定义。

表 2 符号和定义
Table 2 Notations and definitions

符号	定义
k	参与方客户端
w_k	参与方客户端本地模型参数
t	迭代轮数
β^t	浓度参数向量
α^t	权重参数向量
w	全局模型参数
γ	Dirichlet 分布
q	参与方挑选概率
C^t	参与方集合
Δ^t	参与方客户端模型更新
C	梯度裁剪边界值
σ	噪声规模
η	学习率
B_k	本地模型训练中选取的批量大小
S^k	参与方用户数据集

算法 2 给出了 DP-DFL 的描述。DP-DFL 框架分为中心服务器端和参与方客户端。在参与方客户端，主要使用 DP-SGD 进行模型的训练和参数的隐私保护，具体步骤如下。

步骤 1 (第 1 行): 初始化本地参与方客户端模型 w_k^t 。

步骤 2 (第 2 行): 随机抽取样本 B_k 。

步骤 3 (第 3 行): 对样本 B_k 中的每一个数据计算梯度 g_d 。

步骤 4 (第 4 行): 裁剪梯度，限制单个数据对整体的影响。

步骤 5 (第 5 行): 在每一个数据梯度上加入高斯噪声后计算平均梯度，并更新参数 w_k^t 。

步骤 6 (第 7 行): 训练完成后返回模型更新 $\Delta_k^t = w_k^t - w_k^{t-1}$ 。

算法 2 DP-DFL

中心服务器

初始化 w^0 , $\alpha^0 = \gamma(\beta^0)$

1) **for** $t = 1, 2, 3, \dots, T$ **do**

2) **for** $k = 1, 2, 3, \dots, K$ **in parallel do**

3) 以概率 q 挑选参与训练的本轮训练的参与方集合 C^t

4) 中心服务器发送 w^{t-1} 给 C^t

5) $\beta^t = \text{动态聚合权重算法}(\beta^{t-1})$

6) $\alpha^t \leftarrow \gamma(\beta^t)$

7) $\Delta_k^t \leftarrow \text{参与方客户端更新}(k, w^{t-1})$

8) $\Delta^t \leftarrow \frac{\sum_{k \in C^t} \alpha_k^t \Delta_k^t}{q \sum_{k=1}^K \alpha_k^t}$

9) $\Delta^t \leftarrow \Delta^t / \max(1, \frac{\|\Delta^t\|}{C})$

10) $w^t \leftarrow w^{t-1} + \Delta^t + N(0, C^2 \sigma^2)$

11) end

12) return w^t

参与方客户端更新 (k, w^{t-1})

初始化 $w_k^t \leftarrow w^{t-1}$

1) 随机抽选 $I \subseteq \{1, 2, \dots, |S^k|\}$ 共计 B_k 个样本量

2) for $d \in I$ do

3) $g_d = \nabla l(w_k; d)$

4) $g_d \leftarrow g_d / \max\{1, \|g_d\|/C\}$

5) $w_k^t \leftarrow w_k^{t-1} - \frac{\eta}{B_k} (\sum_{d \in I} g_d + N(0, C^2 \sigma^2))$

6) end for

7) return $\Delta_k^t = w_k^t - w_k^{t-1}$

在中心服务器端，每一轮迭代 t 中，中心服务器首先向部分参与方客户端发送全局模型 w^{t-1} 。多个参与方客户端接收到全局模型 w^{t-1} 后采用参与方客户端进行更新，输出本地模型更新参数 Δ_k^t 。执行动态聚合算法，动态聚合权重算法要求每个参与方客户端从自己的本地数据中取样一个小型批次 x ，通过动态聚合权重算法求得一组可学习的权重函数 $\beta^t = [\beta_1^t, \beta_2^t, \dots, \beta_K^t]$ ，这组函数被用于中心服务器的步骤 6 中，通过公式 $\alpha^t = \gamma(\beta^t)$ 计算得到可学习的权重 $\alpha^t = [\alpha_1^t, \alpha_2^t, \dots, \alpha_K^t]$ 。 $\alpha^t = \gamma(\beta^t)$ 函数主要通过 Dirichlet 分布方法求得。得到动态聚合权重后，随机挑选参与训练的参与方客户端集合 C^t ，对挑选的参与方客户端执行本地模型训练。中心服务端接收每一个客户端 k 的模型更新参数 Δ_k^t 并执行聚合操作，得到 Δ^t 。裁剪 Δ^t ，并通过高斯机制 $N(0, C^2 \sigma^2)$ 生成噪声数

据。在全局模型聚合操作中添加噪声数据，得到新的全局模型 w^t 参数。重复多轮迭代，直到达到要求的迭代次数或收敛。

4 实验与结果

4.1 实验环境设置

实验环境：操作系统为 Windows 10，CPU 为 InterCore i5-5250U，基于框架为 TensorFlow2.0，开发环境为 Anaconda3、Python3.8.0、Pycharm、CUDA Toolkit 10.0。

数据集：CIFAR-10 数据集共用 60 000 幅图片，每幅照片均为 32×32 的彩色照片，每个像素点包含 R、G、B 这 3 个数值，数值范围为 0~255，照片可以划分为 10 种不同的类别。其中，50 000 幅图片划为训练集，10 000 幅图片划为测试集。

4.2 结果分析

(1) non-iid 结果分析

本文实验使用两个常用的图像分类问题 CNN 模型 Resnet18 和 VGG16。Resnet18 网络分为 4 层，每层有两个模块组成，第一层由两个普通的残差块组成，其他 3 层由一个普通的残差块和下采样的卷积块组成。VGG16 网络中，主要有 13 层卷积层以及 3 层全连接层，其中卷积层和全连接层使用 ReLU 函数作为激活函数。为了模拟异质数据的环境，对数据进行 non-iid 操作，本文实验使用 Zheng 等^[27]的方法对数据集进行划分。

主要参数设定如下，参与方客户端 k 为 10，参与方客户端本地迭代次数为 2，中心服务器迭代轮数为 10，本地模型训练学习率 η 为 0.01，batchsize 为 32，本地模型训练时的噪声规模 σ_1 和中心服务器聚合时的噪声规模 σ 均为 0.01，本地和中心服务器在进行梯度裁剪时的边界值 C 取 100，挑选每轮参与训练的本轮训练的参与方的概率 q 为 0.2 和 0.5。

图 2 是模型设定为 Resnet18 时的实验结果，其中，图 2(a)挑选每轮参与训练的本轮训练的参与方的概率 q 为 0.2，图 2(b)概率 q 为 0.5。根据迭代 10 次后的模型最终准确率，DP-DFL 优于 DP-FedAvg，图 2(a)中 DP-DFL 准确率为 68.81%，高于 DP-FedAvg 的 2.19%，图 2(b)中 DP-DFL 准确率为 78.56%，高于 DP-FedAvg 的 2.42%。图 3

是模型设定为 VGG16 时的实验结果, 其中图 3(a) 挑选每轮参与训练的本轮训练参与方的概率 q 为 0.2, 图 3(b) 概率 q 为 0.5。根据迭代 10 次后的模型最终准确率, DP-DFL 优于 DP-FedAvg, 图 3(a) 中 DP-DFL 准确率为 81.37%, 高于 DP-FedAvg 的 2.36%, 图 3(b) 中 DP-DFL 准确率为 88.21%, 高于 DP-FedAvg 的 1.39%。

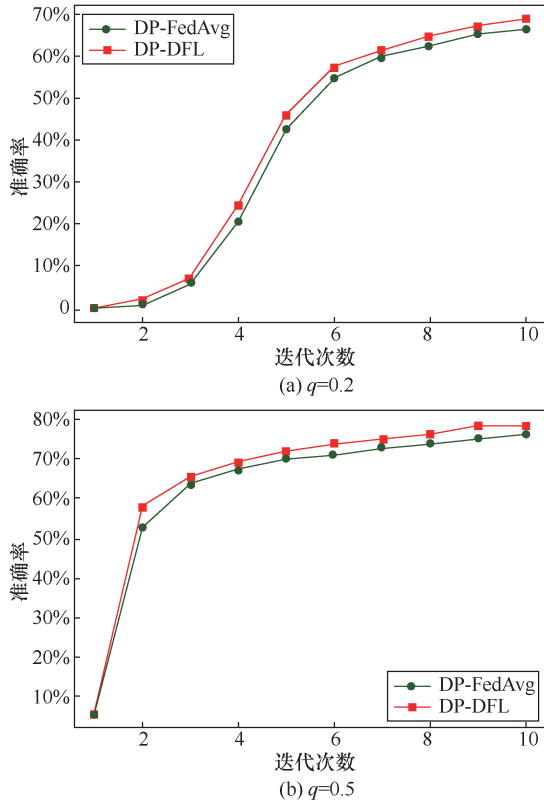


图2 Resnet18 实验结果

Figure 2 The experimental result of Resnet18

根据结果显示, 无论训练模型选择 Resnet18 或 VGG16, 无论每轮迭代选取的本地参与方客户端数量为 2 或 5, DP-DFL 在保障本地隐私的条件下, 不仅模型精度没有受损, 相较 DP-FedAvg 算法的模型准确率平均提高了 2.09%。

差分隐私以损失部分准确率来达到隐私保护的效果。将 DP-DFL 框架与其他两个非 DP 环境下的联邦学习方法相比, 模型准确率也未受到很大的影响。

(2) iid 结果分析

本文实验使用常见的 CNN 模型, 包括 2 个具有 3×3 核的卷积层, 3 个具有 128 个单元和

ReLU 激活函数的全连接层, 以及 1 个最终的输出层。向客户端平均分配 Cifar10 图像, 即处于同分布环境。

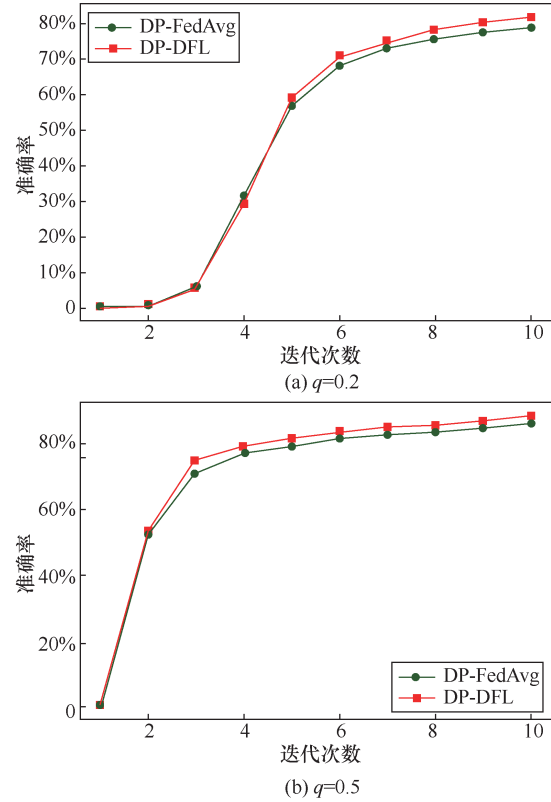


图3 VGG16 实验结果

Figure 3 The experimental result of VGG16

主要参数设定如下: 参与方客户端 k 为 10, 参与方客户端本地迭代次数为 2, 中心服务器迭代轮数为 10, 本地模型训练学习率 η 为 0.01, batch size 为 32, 本地模型训练时的噪声规模 σ_l 和中心服务器聚合时的噪声规模 σ 均为 0.01, 本地和中心服务器在进行梯度裁剪时的边界值 C 取 100, 挑选每轮参与训练的本轮训练的参与方的概率 q 为 0.2 和 0.5。各方法的实验结果如表 3 所示。

图 4 是 iid 环境下的实验结果, 其中图 4(a) 挑选每轮参与训练的本轮训练的参与方的概率 q 为 0.2, 图 4(b) 概率 q 为 0.5。结果表明, 在同分布情况下, DP-DFL 收敛速度虽然慢于 DP-FedAvg, 但在第 10 次全局迭代左右, 两者准确率已经达到相似水平。DP-DFL 在同分布数据情况下仍可以保持良好的性能表现。

表 3 实验结果
Table 3 The experimental result

网络	方法	迭代 10 次后准确率	
		$q=0.2$	$q=0.5$
Resnet 18	DP-FedAvg	66.62%	76.14%
	DP-DFL	68.81%	78.56%
	FedSGD	67.34%	76.98%
	FedAvg	68.84%	78.35%
VGG16	DP-FedAvg	79.01%	86.82%
	DP-DFL	81.37%	88.21%
	FedSGD	80.67%	87.34%
	FedAvg	81.23%	88.28%

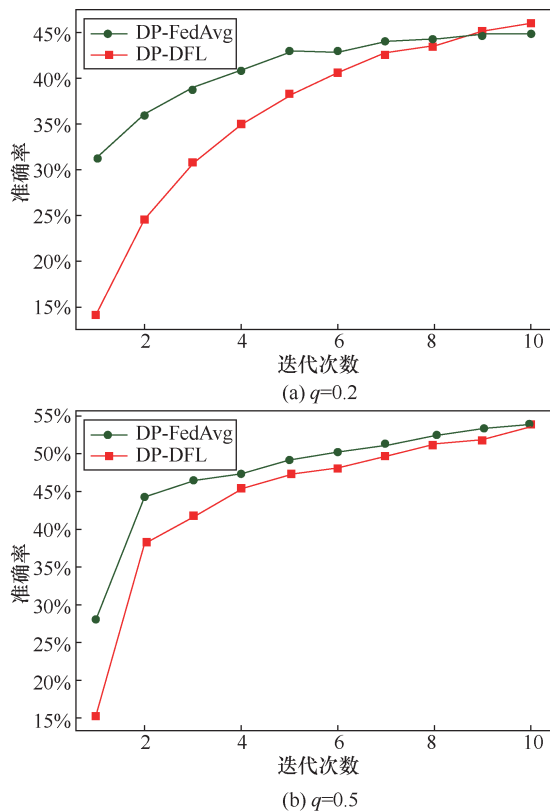


图 4 iid 环境下的实验结果
Figure 4 The experimental result of iid

5 结束语

针对现有联邦学习全局模型聚合时, 不同参与方客户端拥有非独立同分布的数据类型造成模型全局模型效用和适用性效果不高, 以及中心服务器不可信的问题, 本文提出了一种 DP-DFL 框架。该框架在本地参与方模型训练阶段引入了 DP-SGD 以保证本地模型参数的隐私不会被不可

信第三方窃取, 同时, 在本地模型参数聚合阶段设计了一种新的动态模型聚合权重的方法。实验验证, DP-DFL 不仅进一步保证了用户隐私, 而且模型精度优于 DP-FedAvg 算法。未来的工作将研究更复杂的聚合操作, 同时可以模拟现实真实攻击, 寻找效用和隐私保护均衡更优的方法。

参考文献:

- [1] 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.
- [2] XU X H, PENG H, SUN L C, et al. Fedmood: federated learning on mobile health data for mood detection[J]. arXiv preprint arXiv:2102.09342, 2021.
- [3] CHE SICONG, PENG H, SUN L C, et al. Federated multiview learning for private medical data integration and analysis[J]. arXiv preprint arXiv:2105.01603, 2021.
- [4] MC-MAHAN B, MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Artificial Intelligence and Statistics. 2017: 1273-1282.
- [5] LIU Z, GUO J, YANG W, et al. Privacy-preserving aggregation in federated learning: a survey[J]. arXiv preprint arXiv: 2203.17005, 2022.
- [6] RAMASWAMY S, THAKKAR O, MATHEWS R, et al. Training production language models without memorizing user data[J]. arXiv preprint arXiv:2009.10031, 2020.
- [7] LI Z, SHARMA V, MOHANTY S P. Preserving data privacy via federated learning: challenges and solutions[J]. IEEE Consumer Electronics Magazine, 2020, 9(3): 8-16.
- [8] LI T, SAHU A K, ZAHEER M, et al. Federated optimization in heterogeneous networks[J]. Proceedings of Machine Learning and Systems, 2020, 2: 429-450.
- [9] 陈前昕, 毕仁万, 林劼, 等. 支持多数不规则用户的隐私保护联邦学习框架[J]. 网络与信息安全学报, 2022, 8(1): 139-150.
- [10] CHEN Q X, BI R W, LIN J, et al. Privacy-preserving federated learning framework with irregular-majority users[J]. Chinese Journal of Network and Information Security, 2022, 8(1): 139-150.
- [10] 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, 2021, 33(3): 1057-1092.
- [11] 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.
- [11] 王腾, 霍峥, 黄亚鑫, 等. 联邦学习中的隐私保护技术研究综述[J]. 计算机应用, 2022.
- [12] WANG T, HUO Z, HUANG Y X, et al. Survey of privacy-preserving technologies in federated learning [J]. Journal of Computer Applications, 2022.
- [12] FIGURNOV M, MOHAMED S, MNIH A. Implicit reparameterization gradients[C]//Advances in Neural Information Processing Systems. 2018: 441-452.
- [13] PAPERNOT N, ABADI M, ERLINGSSON U F, et al. Semi-supervised knowledge transfer for deep learning from private

- training data[C]//ICLR. 2017.
- [14] SONG C, RISTENPART T, SHMATIKOV V. Machine learning models that remember too much[C]//2017 ACM SIGSAC Conf. on Computer and Communications Security. 2017: 587-601.
- [15] SHOKRI R, STRONATI M, SONG C, et al. Membership inference attacks against machine learning models[C]//2017 IEEE Symposium on Security and Privacy (SP). 2017: 3-18.
- [16] KIM M, GÜNLÜ O, SCHAEFER R F. Federated learning with local differential privacy: Trade-offs between privacy, utility, and communication[C]//ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2021: 2650-2654.
- [17] SUN L, QIAN J, CHEN X. LDP-FL: practical private aggregation in federated learning with local differential privacy[J]. arXiv preprint arXiv:2007.15789, 2020.
- [18] 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.
- [19] MIRONOV I, TALWAR K, ZHANG L. R'enyi differential privacy of the sampled gaussian mechanism[J]. arXiv preprint arXiv:1908.10530, 2019.
- [20] SUN L, LYU L. Federated model distillation with noise-free differential privacy[J]. arXiv preprint arXiv:2009.05537, 2020.
- [21] NAGAR A, TRAN C, FIORETTO F. Privacy-preserving and accountable multi-agent learning[C]//AAMAS Conference proceedings. 2021.
- [22] DAI Z, LOW B K H, JAILLET P. Differentially private federated Bayesian optimization with distributed exploration[J]. Advances in Neural Information Processing Systems, 2021, 34.
- [23] ANDREW G, THAKKAR O, MCMAHAN B, et al. Differentially private learning with adaptive clipping[J]. Advances in Neural Information Processing Systems, 2021, 34.
- [24] MC-MAHAN H B, RAMAGE D, TALWAR K, et al. Learning differentially private recurrent language models[C]//arXiv preprint arXiv:1710.06963, 2017.
- [25] GEYER R C, KLEIN T, NABI M. Differentially private federated learning: a client level perspective[J]. arXiv preprint arXiv:1712.07557, 2017.
- [26] WANG J Y, LIU Q H, LIANG H, et al. Tackling the objective inconsistency problem in heterogeneous federated optimization[C]//NeurIPS. 2020.
- [27] ZHENG Q, CHEN S, LONG Q, et al. Federated f-differential privacy[C]//International Conference on Artificial Intelligence and Statistics. 2021: 2251-2259.
- [28] HU R, GUO Y X, LI H N, et al. Personalized federated learning with differential privacy[J]. IEEE Internet of Things Journal, 7(10):9530-9539, 2020.
- [29] CHEN X N, WANG R C, CHENG M H, et al. Drnas: dirichlet neural architecture search[J]. arXiv preprint arXiv:2006.10355, 2020.
- [30] XIA Y, YANG D, LI W, et al. Auto-FedAvg: learnable federated averaging for multi-institutional medical image segmentation[J]. arXiv preprint arXiv:2104.10195, 2021.
- [31] ZHAO Y, ZHAO J, YANG M, et al. Local differential privacy-based federated learning for internet of things[J]. IEEE Internet of Things Journal, 2022, 8(11):8836-8853.
- [32] LI Y W, CHANG T H, CHI C Y. Secure federated averaging algorithm with differential privacy[C]//2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP), 2020: 1-6.

[作者简介]



应作斌（1982- ），男，安徽芜湖人，澳门城市大学助理教授，主要研究方向为区块链、联邦学习。



方一晨（1998- ），女，浙江湖州人，澳门城市大学硕士生，主要研究方向为差分隐私、联邦学习。



张怡文（1980- ），女，安徽阜阳人，安徽新华学院教授，主要研究方向为数据挖掘、联邦学习。