计算机应用
Journal of Computer Applications
ISSN 1001-9081,CN 51-1307/TP

《计算机应用》网络首发论文

题目：基于个性化差分隐私的联邦学习算法
作者：尹春勇，屈锐
收稿日期：2022-03-18
网络首发日期：2022-06-01
引用格式：尹春勇，屈锐．基于个性化差分隐私的联邦学习算法[J/OL]．计算机应用．
https://kns.cnki.net/kcms/detail/51.1307.TP.20220530.2032.012.html

Journal of Computer Applications
计算机应用
ISSN 1001-9081
CODEN JYIIDU
http://www.joca.cn

DOI:10.11772/j.issn.1001-9081.2022030337

基于个性化差分隐私的联邦学习算法

尹春勇*,屈　锐
(南京信息工程大学 计算机与软件学院,南京 210044)
(*通信作者电子邮箱 yinchunyong@hotmail.com)

摘　要:作为分布式机器学习的一种范式,联邦学习可以有效保护用户的个人数据不被攻击者获得。但是,通过分析模型训练中的参数,如深度神经网络训练的权值,仍然可能泄露用户的隐私信息。为了解决这个问题,差分隐私被应用到联邦学习中去实现联邦学习的隐私增强。然而,现有的联邦学习差分隐私方法只关注统一的隐私保护预算,而忽略了用户的个性化隐私需求。针对此问题,提出了一种两阶段的基于个性化差分隐私的联邦学习(PDP-FL)算法。在第一阶段,依据用户的隐私偏好对用户隐私进行分级,并添加满足用户隐私偏好的噪声,达到个性化隐私保护的目的,同时上传隐私偏好对应的隐私等级给中央聚合服务器。在第二阶段,为实现对全局数据的充分保护,采取本地和中心同时保护的策略,根据用户上传的隐私等级,来添加符合全局差分隐私阈值的噪声,量化了全局的隐私保护水平。实验结果表明,相比基于本地化差分隐私的联邦学习(LDP-Fed)方法,PDP-FL在实现多个场景下的分类准确度提高了0.9%的同时,达成了个性化隐私保护的需求。

关键词:联邦学习;差分隐私;隐私偏好;隐私分级;个性化隐私保护
中图分类号:TP309.2　　**文献标志码**:A

Federated learning algorithm based on personalized differential privacy

YIN Chunyong*, QU Rui
(*School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing* 210044, *China*)

Abstract: As a paradigm of distributed machine learning, federated learning can effectively protect users′ personal data from attackers. However, by analyzing the parameters in the model training, such as the weight of deep neural network training, it was still possible to disclose the user′s privacy information. In order to solve this problem, differential privacy was applied to federated learning to enhance the privacy of federated learning. However, existing federated learning methods based on differential privacy revolved around the same privacy protection budget and ignore the different privacy requirements of clients. To solve this problem, a two-stage Federated Learning algorithm based on Personalized Differential Privacy (PDP-FL) was proposed. In the first stage, the user′s privacy was graded according to the user′s privacy preference, and the noise meeting the user′s privacy preference was added to achieve the purpose of personalized privacy protection. At the same time, the privacy level corresponding to the privacy preference was uploaded to the central aggregation server. In the second stage, in order to fully protect the global data, the local and central protection strategy was adopted. According to the privacy level uploaded by the user, the noise conforming to the global differential privacy threshold was added to quantify the global privacy protection level. The experimental results showed that compared with federated learning with local differential privacy (LDP-Fed) algorithm, PDP-FL algorithm not only improveed the classification accuracy in multiple scenes by 0.9%, but also meetedthe needs of personalized privacy protection.

Key words: federated learning; differential privacy; privacy preference; privacy rating; personalized privacy protection

0　引言

随着大数据驱动的人工智能的快速发展,在医疗保健、食品农业、智能交通等生活方面出现了大量的有关应用[1-2]。然而,机器学习也面临着新的困难和挑战。一方面,机器学习希望为所有用户提供健壮、高效的服务。另一方面,数据作为机器学习算法的基础,难以完全共享[3-4]。例如,在互联网场景中,大部分企业不愿意分享用户的原始数据。为解决这一问题,联邦学习(Federated Learning, FL)应运而生[5-6]。FL的主要创新是在大量参与者的合作下提供有效的隐私保护,以分布式计算的方式迭代训练特定的机器学习模型。FL为机器学习提供了一个双赢的前景,在保证数据隐私的同时提高了机器学习的性能。在FL中,用户的原始数据不需要上传到中央聚合服务器,在共享每个参与用户本地训练模型的同时,保证每个参与者的数据安全[7]。

然而,现有的研究结果表明,FL也存在一些安全问题。例如,攻击者对客户端训练上传的相关参数的查询差异进行分析,可能会获取参与用户的部分隐私信息[8]。尤其是内部实体发起的攻击(如恶意用户、不信任的服务器等)的威胁更大[9]。针对这一问题,学者提出了各种解决方案,如基于同态加密的联邦学习和基于安全多方计

收稿日期:2022-03-18;**修回日期**:2022-05-13;**录用日期**:2022-05-20。
作者简介:尹春勇(1977—),男,山东潍坊人,教授,博士,博士生导师,主要研究方向:网络空间安全、大数据挖掘、隐私保护、人工智能、新型计算;屈锐(1999—),男,江苏宿迁人,硕士研究生,主要研究方向:差分隐私、联邦学习。

算的联邦学习等，然而这些方法的计算和通讯代价很高[10-12]。基于差分隐私（Differential Privacy，DP）的联邦学习算法与其他方法相比，具有明显的轻量级优势[13]。目前，基于差分隐私技术的联邦学习工作主要包括两类：1）用户在上传参数之前，利用本地差分隐私来扰动用户上传的模型参数[14]；2）利用中心化差分隐私对中央聚合服务器的汇聚梯度进行扰动[15-16]。

然而，用户的隐私需求会因职业和地域的不同而有所不同。上述的大部分工作是基于平均分配的隐私预算，由于法律、国家或职业背景不同，这种假设是不切实际的。此外，统一的隐私预算级别意味着一些客户会浪费大量的隐私预算，这往往会对模型的准确性产生负面影响[17]。进一步，在个性化的联邦学习中，用户参差不齐的隐私偏好会降低模型的训练效率，并且不能量化隐私保护程度。

为了解决上述问题，提出了一种两阶段的基于个性化差分隐私的联邦学习（Federated Learning with Personalized Differential Privacy，PDP-FL）算法。用户可以在本地设置他们的隐私偏好，然后个性化差分隐私算法根据用户的隐私偏好来添加相应的噪声来降低隐私预算的浪费。算法通过设置需要满足的全局差分隐私的隐私预算阈值来控制添加噪声的大小，量化了隐私保护程度的同时，对联邦学习进行全局保护。首次实现了在满足用户的个性化隐私需求的情况下，同时保护了本地和中心的隐私，并量化了全局隐私保护强度。主要贡献有四点：

1）提出了一种基于个性化差分隐私的联邦学习方法。算法可以根据用户设置的隐私偏好来保护信息，达到个性化隐私保护的目的。

2）提出了一种两阶段的联邦学习本地和中心同时保护算法，实现了对噪声添加尺度的控制和量化全局隐私保护的目标，解决了用户个性化的隐私偏好对模型训练效率的影响。

3）针对部分用户添加过量噪声影响模型训练效果的情况，算法根据用户的隐私偏好设定了隐私分级。隐私分级一方面控制了噪声的最大尺度，另一方面中央聚合服务器通过隐私等级（低，中，高）来判断用户的模型参数的聚合权重。

4）实验评估了在不同的情景模式下，不同隐私预算分布需求的用户对PDP-FL算法的成功率的影响。

1 相关工作

针对联邦学习的安全问题，同态加密、安全多方计算和差分隐私等技术被使用去保护联邦学习过程的安全。Aono等[18]提出了使用加法同态机制对联邦学习过程进行保护，通过对上传的梯度进行加密，来保护用户的隐私安全。Song等[19]在秘密共享的基础上，设计了一种高效的保护隐私的FL数据聚合机制，以抵抗反向攻击。Gong等[20]在多方安全计算的基础上，动态分配隐私预算，提高了模型的准确度。然而上诉的方法的计算和通信开销较大，与此相比，差分隐私具有轻量级的优势，比较适合联邦学习的应用场景。

基于差分隐私的联邦学习算法主要分为面向用户端的本地化差分隐私联邦学习和面向中央聚合服务器的中心化差分隐私联邦学习[21-24]。在面向中央聚合服务器的中心化差分隐私联邦学习中，Wang等[25]提出了一种基于差分隐私的随机梯度下降算法的解决方案，这是各种机器学习任务的基础。由于记录级别差分隐私无法抵抗属性推理攻击，Geyer等[26]首次提出了在联邦学习下针对用户级别的差分隐私保护算法。用户级别的隐私保护是指相邻数据集相差的是一个用户所有的记录。针对中央聚合服务器的不可信问题，Truex等[27]将本地化差分隐私引入了联邦学习中，通过本地差分隐私模块对用户上传的梯度的扰动来保护用户的隐私。针对服务器聚合后会稀释噪声，Wei等[28]设计了分阶段差分隐私保护模型。通过建立全局的差分隐私定义，对联邦学习中的上传和广播信道都进行了保护。为了在传输速率和隐私的约束下最大化用户的收敛速度，Wu等[29]将瑞丽差分隐私应用到联邦学习中，提高了联邦学习的收敛速度。面向用户端的本地化差分隐私联邦学习中，针对上传权重维数过多的问题，Liu等[30]提出了一种改进的权值传递方案，通过指数机制选择权重最高的k个维度进行扰动。针对联邦学习下的多源异构数据融合的问题，莫慧凌等[31]提出了基于联邦学习的多源异构数据融合算法，通过对异构数据进行融合，使对多源异构的个人数据进行个性化处理提供了方式。Zhao等[32]在本地和中心分别设计了隐私保护算法，使得整个联邦学习训练和模型结果受到$Max(\varepsilon_1,\varepsilon_2)$-差分隐私保护。为了克服本地化差分隐私对联邦学习的性能影响，Girgis等[33]将shuffled隐私模型运用于联邦学习，提高了联邦学习的通信效率。针对联邦学习的用户参与问题，Zhang等[34]通过引入博弈论来激励用户参与到联邦学习中。

然而，这些工作都没有考虑到用户的个性化需求。一方面，由于背景环境的不同，统一的隐私预算难于在现实中实现。另一方面，浪费大量的隐私预算，通常会对模型的准确性带来负面影响。Avent等[35]考虑了普通用户和信任服务器用户的组合，提出了混合差分隐私模型，针对普通用户，采用本地化差分隐私对数据进行保护，而对于信任服务器的用户，可以降低保护程度或者直接发送原始参数。但面对多种隐私需求的参与用户，Avent等提出的算法难以兼顾。Hu等[36]提出了一种个性化联邦学习方案，通过学习用户的特征来完成多任务学习。但Hu等没有考虑联邦学习过程的安全性。Yang等[37]提出了一种个性化的差分隐私联邦学习方案，在该方案中，用户通过使用自己的隐私偏好来扰动数据。但Yang等的工作没有考虑到用户个性化添加噪声对于全局数据的影响，也没有量化隐私保护程度。

与上述方案不同，本文提出了PDP-FL算法。一方面，PDP-FL在本地根据用户设置的隐私偏好对用户数据进行保护，满足了用户的个性化需求。另一方面，通过引入全局差分隐私的概念，解决了用户个性化的隐私偏好对模型训练效率的影响，量化了隐私保护程度。综上所述，提出的方案实现了在满足用户的个性化隐私需求的同时，降低了个性化隐私偏好对成功率的影响。

2 预备知识

2.1 差分隐私

DP为联邦学习的隐私保护提供了一个强有力的准则，通过在模型参数上添加满足差分隐私的噪声，来实现联邦学习的隐私增强。以下是有关差分隐私的一些定义[38-39]：

定义1 差分隐私：随机算法$M: X \to R$，X是定义域，R是值域。若随机算法M在任意相邻数据集D和D'上满足公式1，则随机算法M满足(ε,δ)-DP。特别当$\delta=0$时，算法M满足ε-差分隐私。

$$\Pr(M(D)\in S)\leqslant e^{\varepsilon}\times\Pr(M(D')\in S)+\delta \tag{1}$$

其中参数ε是隐私保护预算，ε的值决定了隐私保护的程度。ε越小，隐私保密性越高；ε越大，数据的可用性越高。

定义2 本地化差分隐私（local differential privacy，LDP）：如果随机算法M在任意两条记录t和t'下输出相同的结果t^*，即满足公式2，则称随机算法M满足ε-LDP。

$$\Pr(M(t)=t^*)\leqslant e^{\varepsilon}\times\Pr(M(t')=t^*) \tag{2}$$

其中参数ε是隐私保护预算。LDP通过保证任意两条记录的输出相似度来保护用户的隐私。

定义3 高斯机制：假设有一个函数$f: D\to R^D$，灵敏度为Δf。当$N\sim N(0,c\Delta f/\varepsilon)$，随机算法$M=f(D)+N$提供了$(\varepsilon,\delta)$-DP的隐私保护。其中参数$c\geq\sqrt{2ln(1.25/\delta)}$。高斯分布的概率密度函数公式如式（3）：

$$P(x)=\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{3}$$

其中σ代表高斯分布的标准差和μ代表高斯分布的期望值，敏感度$\Delta f=\max_{D,D'}\|f(D)-f(D')\|$，$f$是查询函数。高斯机制采用的是$l_2$灵敏度，比较适合联邦学习的训练过程。实际上其他的差分隐私实现机

制也可以完成加噪的任务,但是为了量化全局隐私保护,需要在两阶段采取相同的噪声添加机制。

2.2 联邦学习

联邦学习的过程可以定义如下:有N个参与者,D_i代表用户F_i持有的本地数据集,其中$i \in \{1, 2, ..., N\}$。在联邦学习中,中央聚合服务器会选择部分用户作为参与用户,参与用户需要合作训练全局模型,而不将本地数据D_i暴露给第三方。本地训练时,客户端需要找到一个模型权重$\boldsymbol{w}_i$,使损失函数最小化。形式上,中央聚合服务器将N个参与者发送的权重$\boldsymbol{w}_i$聚合公式如下:

$$\boldsymbol{w} = \sum_{i=1}^{N} p_i \boldsymbol{w}_i \tag{4}$$

其中,$\boldsymbol{w}_i$为第i个参与用户上传的模型参数,$\boldsymbol{w}$是聚合后的模型参数,N是参与用户的数量,p_i是第i个参与用户数据集占数据总量的比例,用于中央服务器的聚合计算。最小化损失函数的过程的公式如下:

$$\boldsymbol{w}^* = \arg\min_{\boldsymbol{w}} \sum_{i=1}^{N} p_i \| m_i(\boldsymbol{w}) - Y \|_p \tag{5}$$

其中:$m_i(\boldsymbol{w})$为函数的输出结果,Y为正确输出值。

2.3 隐私偏好

隐私偏好是指用户对个人隐私信息的重视程度和个人隐私数据的保护倾向,即对个人隐私信息暴露的可接受性。用户对敏感属性数据有个性化的隐私要求,而传统的隐私保护方案采用统一的隐私保护级别,不能很好地保护用户的隐私。为了定量分析隐私保护过程中的隐私泄露,通过数学方法对用户的隐私偏好进行形式化的定义,从而帮助用户表达自己的隐私偏好。

定义4 隐私偏好:将隐私偏好等级分为$1, 2, ..., r$。n个用户的隐私偏好集为$R = \{R_1, R_2, ..., R_i, ..., R_n\}$,$R_i$代表用户根据隐私需求对敏感数据的隐私偏好评估等级,用户可以在本地选择隐私偏好集的隐私偏好。

因为每个用户对数据的敏感性是不确定的,可以基于特定的隐私偏好程度为参与用户考虑特定级别的隐私保护。在参与用户将其隐私数据发送到服务器之前,需要一些本地隐私机制来根据用户的隐私偏好来干扰数据。

定义5 个性化差分隐私(personalized differential local privacy, PDP)[40-41]:对于ε_i,由用户的隐私偏好与总隐私预算决定,根据其值的大小决定本地噪声的添加。对于用户i,如果随机算法M在任意两条记录t和t'下输出相同的结果t^*,即满足公式6,则称M满足ε-PDP。

$$\Pr(M_i(t) = t^*) \leqslant e^{\varepsilon_i} \times \Pr(M_i(t') = t^*) \tag{6}$$

用户i的隐私预算ε_i由用户隐私偏好R_i得到。借鉴了文献[41]提出的方法,将用户的隐私偏好通过公式7计算出隐私预算,来生成符合用户隐私偏好的噪声。假设给定的隐私预算值为ε,则用户i的隐私预算计算如:

$$\varepsilon_i = R_i \log_2 \varepsilon \tag{7}$$

隐私预算按照隐私预算等级分类策略$\varepsilon(\varepsilon_{low}, \varepsilon_{mid}, \varepsilon_{high})$划分为三个等级,判断用户$i$的隐私偏好$R_i$属于哪一等级,将该等级的值赋予隐私保护等级$\varepsilon_k'$,用于中央服务器添加第二阶段噪声。具体见算法2中的算法描述部分。

3 基于个性化差分隐私的联邦学习

在本部分,首先给出了问题的定义,即提出的PDP-FL算法需要解决的问题。其次,从参与客户端选择和参数广播、PDP-FL本地更新和PDP-FL联邦聚合三个部分来介绍提出的PDP-FL算法,并说明了算法在威胁模型下的高鲁棒性。最后,给出了在用户个性化设置隐私参数的情况下,中央聚合服务器对数据处理的相关参数推导并证明。

3.1 问题定义

联邦学习的过程大致分为四个阶段:1)本地训练:所有的参与用户在本地训练出模型参数,根据用户的隐私偏好添加高斯噪声,并将模型参数传输到中央聚合服务器;2)模型聚合:服务器聚合参与用户上传的参数,判断客户端添加的噪声是否符合全局差分隐私的要求,若不满足,则继续添加噪声以满足全局差分隐私;3)参数广播:服务器将参数广播给下一轮参与的客户端;4)模型更新:参与用户使用全局参数更新本地模型,开始下一轮的本地训练。

联邦学习隐私保护的主要目标是在保护参与用户的信息的情况下,尽量不损失模型的精度。由此进行了三个方向的考虑:1)提供个性化的隐私保护,让参与用户自主决定隐私保护的级别。2)提供本地和中心同时保护的功能,避免被恶意用户和不可信任的服务器获取用户的信息。3)保证模型训练的正常进行,用户不能在本地添加影响模型正常训练的噪声,要保证模型训练的正常进行。基于个性化差分隐私的联邦学习框架如图1所示。

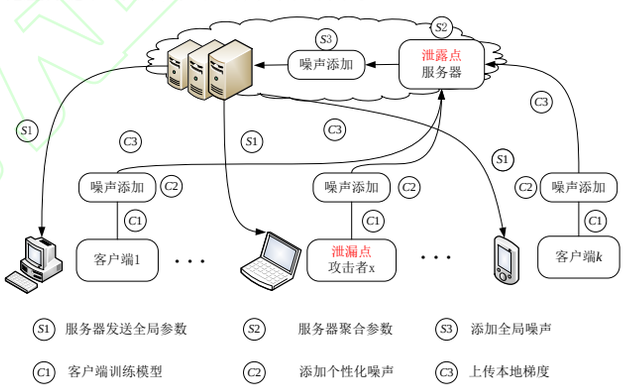


图1 基于个性化差分隐私的联邦学习框架

Fig. 1 Federated learning systems based on personalized differential privacy

为了达到联邦学习隐私保护的目标,对比了相关联邦学习的方法,通过对同态加密、多方安全计算和差分隐私等技术的对比,最终选择个性化差分隐私来实现联邦学习隐私保护的目标。具体的对比方案见表1。

表1 基于联邦学习的隐私保护方法比较

Tab. 1 Comparison of Privacy Protection Methods Based on Federated Learning

保护方法	优点	缺点	保护对象	保护场景
同态加密[18]	准确性高、隐私保护严格。	计算开销高、通信开销高、不保护发布模型。	本地梯度	统一预算场景
多方安全计算[19][20]	准确性较高、隐私保护严格。	通信开销高、协议复杂脆弱	本地梯度	统一预算场景
中心化差分隐私[25][26]	准确性较高、通信开销低、计算量较低。	需要中心服务器可信。	中心梯度	统一预算场景
本地化差分隐私[27]	通信开销低、扰动方案灵活。	准确性低。	本地梯度	统一预算场景
安全混洗[33]	准确性较高、隐私保护严格。	计算开销高。	本地梯度	统一预算场景
PDP-FL算法	可以提供个性化隐私保证,准确性较高。	需要合理的隐私分级。	本地和中心梯度	个性化预算场景

3.2 提出的模型

从三个阶段介绍了提出的PDP-FL算法。具体三个阶段如下:参与用户的选择和参数广播、PDP-FL本地更新和PDP-FL中央聚合。

(1) 参与用户的选择和参数广播:PDP-FL算法分为PDP-FL本

地更新和PDP-FL中央聚合两部分，通过本地训练和中央聚合的不断交互，最终获得训练完成的全局模型参数$\boldsymbol{w}^T$。在参与客户端选择和参数广播阶段，中央聚合服务器随机选择k个客户端后（第1行），向客户端提交全局模型参数和隐私分类策略（第2行）。之后进行PDP-FL本地更新算法（第3行）和PDP-FL中央聚合算法（第4行）的交互。伪代码在算法1中描述如下：

算法1 PDP-FL

输入 初始化的模型参数$\boldsymbol{w}^0$；聚合轮次T；学习率μ；客户端的数量N；每一轮选择的参与用户数量k

输出 全局模型参数$\boldsymbol{w}^T$

1）for 轮次t=1，2，…，T：

2）随机选择k个客户端，并广播$\boldsymbol{w}^t$和$\varepsilon(\varepsilon_{low},\varepsilon_{mid},\varepsilon_{high})$给选择的客户端

3）$\tilde{\boldsymbol{w}}_{t+1}^k,\varepsilon_k$ = PDP - FL本地更新$(\boldsymbol{w}_t,k)$

4）$\boldsymbol{w}_{t+1}$ = PDP - FL中央聚合$(\tilde{\boldsymbol{w}}_{t+1}^k,\varepsilon_k)$

5）return $\boldsymbol{w}^T$

（2）PDP-FL本地更新：PDP-FL本地更新算法根据服务器广播的初始全局参数$\boldsymbol{w}_t$和参与用户的隐私偏好R_k对模型进行训练，将训练完成的模型参数$\boldsymbol{w}_{t+1}^k$和隐私保护等级ε_k'上传到中央聚合服务器。在PDP-FL本地更新阶段，每个客户端根据服务器和本地数据集发送的全局梯度计算本地梯度（第2-5行）。同时，客户端根据参与用户的隐私偏好选择隐私级别（第6-11行），并初始化个性化差分隐私模型，扰动需要上传的梯度（第13行）。敏感度由$\boldsymbol{w}_i$的裁减阈值决定（第12行），隐私预算是根据参与用户的个性化隐私偏好生成的。最后，将扰动后的模型权重$\tilde{\boldsymbol{w}}_{t+1}^k$和隐私保护等级$\varepsilon_k'$上传到服务器（第14行）。PDP-FL本地更新算法的伪代码在算法2中描述如下：

算法2 PDP-FL本地更新

输入 隐私偏好R_k；学习速率μ_k；全局模型参数$\boldsymbol{w}_t$；隐私等级分类策略$\varepsilon(\varepsilon_{low},\varepsilon_{mid},\varepsilon_{high})$

输出 扰动后的本地模型权重$\boldsymbol{w}_{t+1}^k$；隐私保护等级ε_k'

1）for 用户k=1，2，…，N：

2）$\boldsymbol{w}_t^k=\boldsymbol{w}_t$

3）for 本地轮次i=1，2，…，E：

4）for 批次$d\in D$：

5）$\boldsymbol{w}_{t+1}^k=\boldsymbol{w}_t^k-\mu\nabla L(\boldsymbol{w}_t^k;\ d_t^k)$

6）if $R_k<\varepsilon_{low}$：

7）$\varepsilon_k'=\varepsilon_{low};\varepsilon_k=R_k\varepsilon_{low}$

8）else if $R_k>\varepsilon_{high}$：

9）$\varepsilon_k'=\varepsilon_{high};\varepsilon_k=R_k\varepsilon_{high}$

10）else：

11）$\varepsilon_k'=\varepsilon_{mid};\varepsilon_k=R_k\varepsilon_{mid}$

12）$\boldsymbol{w}_{t+1}^k=\boldsymbol{w}_{t+1}^k/max(1,\frac{\|\boldsymbol{w}_{t+1}^k\|}{C})$

13）$\tilde{\boldsymbol{w}}_{t+1}^k=\boldsymbol{w}_{t+1}^k+N(\Delta s_U^{D_k}/\varepsilon_k)$

14）return $\tilde{\boldsymbol{w}}_{t+1}^k,\varepsilon_k'$

（3）PDP-FL中央聚合：PDP-FL中央聚合算法根据参与用户上传的模型参数$\tilde{\boldsymbol{w}}_{t+1}^k$和参与用户的隐私偏好$\varepsilon_k'$进行聚合，并根据是否满足全局差分隐私进行添加噪声，最后将聚合完成的模型参数$\boldsymbol{w}_{t+1}$广播给下一轮的参与用户。在PDP-FL中央聚合阶段，服务器接收到客户端上传的本地模型权重和隐私等级后进行聚合，然后更新一轮的参数（第1-3行）。在个性化联邦学习隐私保护的场景下，用户添加的噪声不统一，为了量化联邦学习过程的差分隐私保护水平，算法需要在中央聚合服务器端来添加噪声，来控制联邦学习的隐私保护水平。如果聚合结果大于全局差分隐私预算阈值，则加入满足该预算阈值的噪声（第7行）。否则，聚合参数将直接发送（第5行）。PDP-FL中央聚合算法的伪代码在算法3中描述：

算法3 PDP-FL中央聚合

输入 客户端上传的模型参数$\tilde{\boldsymbol{w}}_{t+1}^k$；隐私保护等级$\varepsilon_k'$

输出 新一轮的全局模型参数$\boldsymbol{w}_{t+1}$

1）$\theta_k=\dfrac{\varepsilon_k'}{\sum_{k=1}^{n}\varepsilon_k'}$

2）$\varepsilon=\sum_{k=1}^{n}\varepsilon_k'$

3）$\tilde{\boldsymbol{w}}_{t+1}=\sum_{k=1}^{N}\tilde{\boldsymbol{w}}_{t+1}^k\theta_k p_i$

4）if $\varepsilon<\varepsilon_{max}$：

5）$\boldsymbol{w}_{t+1}=\tilde{\boldsymbol{w}}_{t+1}$

6）else：

7）$\boldsymbol{w}_{t+1}=\tilde{\boldsymbol{w}}_{t+1}+N(0,\dfrac{2cC\sqrt{\frac{1}{\varepsilon_{max}^2}-\frac{1}{\varepsilon^2}}}{mN})$

8）return $\boldsymbol{w}_{t+1}$

3.3 威胁模型

目前，联邦学习主要有三个漏洞点：中央聚合服务器、其他参与者和模型本身，分别对应着上传信道、广播信道和全局信道三个泄露的途径。参与用户的信息可能通过上传信道泄漏到中央聚合器服务（泄漏点1）以及通过广播信道发送模型参数泄露给其他参与者（泄漏点2）。此外，最终模型本身也会泄漏预测输出（泄漏点3），导致攻击者推断出关于训练数据的信息。从这三个方面来分析提出的PDP-FL算法的高鲁棒性：

针对中央聚合服务器的泄露问题，PDP-FL算法要求用户在本地训练完自己的参数后，经过差分隐私扰动后再上传到服务器。这避免了被中央聚合服务器直接获取目标用户的信息，保护了上传信道的安全。

针对其他参与者的泄露问题，PDP-FL算法要求中央聚合服务器获取参与用户的信息后，添加符合全局差分隐私的噪声后再发送给新一轮的参与用户。这可以有效避免其他用户根据全局梯度来获取目标用户的信息，保护了广播信道的安全。

针对模型本身的泄露问题，PDP-FL算法符合$(\varepsilon_{max},\delta)$-全局差分隐私，量化了个性化隐私保护水平。因为不需要考虑攻击者的背景知识，输出模型本身有较好的隐私保护水平。

为了证明提出的PDP-FL算法满足$(\varepsilon_{max},\delta)$-全局差分隐私，给出了相应的证明，过程如下：

$$\left|\ln\frac{Pr(M(D_i)=o_i)}{Pr(M(D_i^i)=o_i)}\right|=\left|\ln\frac{Pr(N=o_i-M(D_i))}{Pr(N=o_i-M(D_i^i))}\right|=\left|\ln\frac{e^{-\frac{x^2}{2\sigma_A^2}}}{e^{-\frac{(x+\Delta s_D)^2}{2\sigma_A^2}}}\right|=\left|\frac{1}{2\sigma_A^2}(2x\Delta s_D+\Delta s_D^2)\right|\leqslant\varepsilon \tag{8}$$

可以得到$x<\dfrac{\sigma_A^2\varepsilon}{\Delta s_D}-\dfrac{\Delta s_D}{2}$，设$\dfrac{\sigma_A^2\varepsilon}{\Delta s_D}-\dfrac{\Delta s_D}{2}=\eta$即需证明：

$$Pr(|x|\geqslant\eta)<\delta\rightarrow Pr(x\geqslant\eta)<\frac{\delta}{2} \tag{9}$$

又因为$Pr(x\geqslant\eta)$的边界如下：

$$Pr(x>\eta)=\int_{\eta}^{\infty}\frac{1}{\sqrt{2\pi}\,\sigma_A}e^{-\frac{x^2}{2\sigma_A^2}}dx\leqslant\frac{\sigma_A}{\sqrt{2\pi}\,\eta}e^{-\frac{\eta^2}{2\sigma_A^2}} \tag{10}$$

即证明：

$$\frac{\sigma_A}{\sqrt{2\pi}\,\eta}e^{-\frac{\eta^2}{2\sigma_A^2}}<\frac{\delta}{2}\Leftrightarrow\ln\left(\frac{\eta}{\sigma_A}\right)+\frac{\eta^2}{2\sigma_A^2}>\ln\left(\sqrt{\frac{2}{\pi}}\frac{1}{\delta}\right) \tag{11}$$

当参与用户的隐私规模达到预定的阈值时，将$\sigma_A=\sigma_U=\dfrac{c\Delta s}{\varepsilon}$代

入公式11,可得$c^2 \geqslant 2\ln(\frac{1.25}{\delta})$,即证明完毕。

当参与用户的隐私规模未达到预定的阈值时,将$\sigma_A = \sqrt{\sigma_D^2 + \overline{\sigma_U^2}} = \frac{c\Delta s_D}{N\varepsilon_{max}}$代入公式11,可得$c^2 \geqslant 2\ln(\frac{1.25}{\delta})$,即证明结束。

3.4 理论分析

为了解决中央聚合服务器对用户上传的隐私保护程度不同的个性化参数如何进行处理的问题,在文献[24]的启发下,提出了在个性化差分隐私需求下的全局差分隐私。在本节中,从上传信道、广播信道和全局信道三个部分来解释所提出的方法的细节和证明。

上传通道:PDP-FL算法根据本地设置中的隐私偏好来初始化个性化差分隐私模块,添加符合用户隐私需求的高斯噪声。最后,将扰动后的模型参数和用户的隐私偏好上传给服务器。首先定义全局(ε,δ)对于上行信道和广播信道的个性化隐私要求。然后,通过裁剪技术,可以保证上行通道中的模型权重$\boldsymbol{w}_i \leq C$,其中$\boldsymbol{w}_i$表示第i个参与用户的训练参数,C为$\boldsymbol{w}_i$的裁剪阈值。本地训练过程公式化如下:

$$s_U^{D_i} = \boldsymbol{w}_i = \arg\min_{\boldsymbol{w}} M_i(\boldsymbol{w}_i, D_i)n = \frac{1}{|D_i|}\sum_{j=1}^{|D_i|}\arg\min_{\boldsymbol{w}} \|m_i(\boldsymbol{w}, D_{i,j}) - Y\|_p \tag{12}$$

其中$D_{i,j}$是D_i中的第j个样本。因此,$s_U^{D_i}$的灵敏度可以表示为:

$$\Delta s_U^{D_i} = \max_{D_i, D_i'} \|s_U^{D_i} - s_U^{D_i'}\| = \max_{D_i, D_i'} \|\frac{1}{|D_i|}\sum_{j=1}^{|D_i|}\arg\min_{\boldsymbol{w}} \|m_i(\boldsymbol{w}, D_{i,j}) - Y\|_p - \frac{1}{|D_i'|}\sum_{j=1}^{|D_i'|}\arg\min_{\boldsymbol{w}} \|m_i(\boldsymbol{w}, D_{i,j}') - Y\|_p\| = \frac{2C}{|D_i|} \tag{13}$$

其中D_i是D_i的相邻数据集,$D_{i,j}'$是D_i'中的第j个样本。综上可得,上行通道的灵敏度可以表示为:

$$\Delta s_U = max\{\Delta s_U^{D_i}\} \tag{14}$$

为了使上行信道满足(ε,δ)-DP,采用高斯机制添加噪声。假设最小的本地数据集为m, C是权值的裁剪阈值。具体的噪声添加如下:

$$\sigma_U = \frac{c\Delta s}{\varepsilon_k}, c \geq \sqrt{2ln(1.25/\delta)}, \Delta s = \frac{2C}{m} \tag{15}$$

广播信道:服务器根据客户端上传的模型参数和隐私等级进行聚合。聚合完成后,得到新一轮的模型参数。从广播信道的角度,服务器的聚合过程如下所示。

$$\boldsymbol{w} = p_1\boldsymbol{w}_1 + p_2\boldsymbol{w}_2 + ... + p_i\boldsymbol{w}_i + ... + p_k\boldsymbol{w}_k \text{ ,其中} 1 \leq i \leq N \tag{16}$$

因此,推导出广播信道的灵敏度为:

$$\Delta s_D^{D_i} = \max_{D_i, D_i'} \|p_i\boldsymbol{w}_i(D_i) - p_i\boldsymbol{w}_i(D_i')\| = p_i s_U^{D_i} \leq 2Cp_i/m \tag{17}$$

全局信道:服务器聚合参与用户的隐私保护规模,来决定是否还需要在中央服务器进行聚合操作。如果参与用户的隐私规模达到了设定的阈值,为了保证模型训练的质量,不再继续进行添加噪声。如果参与用户的隐私规模没有达到设定的阈值,需要在中央服务器添加新一轮的噪声。由此,得到的全局信道敏感度如下:

$$\Delta s_D = max\{\Delta s_D^{D_i}\} \tag{18}$$

为了量化结果,假设客户端使用相同大小的数据集,即$p_i = \frac{1}{N}$。因此,全局应当添加的噪声大小如下:

$$\sigma_A = \frac{c\Delta s_D}{\varepsilon}, c \geq \sqrt{2ln(1.25/\delta)}, \Delta s_D = \frac{2C}{mN} \tag{19}$$

因此,在未达到全局差分隐私预算阈值的时候,中央聚合服务器应当添加的噪声为:

$$\sigma_D = \sqrt{\sigma_A^2 - \overline{\sigma_U^2}} = \frac{2cC\sqrt{\frac{1}{\varepsilon_{max}^2} - \frac{1}{\varepsilon^2}}}{mN} \tag{20}$$

4 实验

在本节中,使用卷积神经网络(convolutional neural networks, CNN)对手写数字数据集(MNIST)[42]和图像识别数据集(CIFAR-10)[43]来测量PDP-FL算法。从不同情境下的隐私预算对模型成功率的影响、不同参数的影响以及PDP-FL与现有方法的比较三个方面验证了所提出的PDP-FL算法的合理性。

4.1 实验设置

在MNIST和CIFAR-10数据集上进行了图像识别实验。MNIST数据集包含6万个训练样本和1万个测试样本。每个样本是一个28×28大小的灰度图像。CIFAR-10数据集包含6万个32×32大小的彩色图像,分为5万个训练样本和1万个测试样本。为了模拟FL场景,将数据集按N个客户端划分。使用CNN全局模型和交叉熵损失函数,模型有2个隐藏层、2个输出层和10个softmax输出层。每个隐藏层具有200个单位,使用Relu激活。实验在NVIDIA GTX 1050 4GB、640 CUDA核心和16 GB内存的Windows 10环境下运行,所有算法采用python 3.8编程语言和pytorch1.7框架实现。

4.2 不同隐私预算场景下的模型性能

由于隐私预算的高低对模型的准确性有很大的影响,为了比较不同隐私预算场景下的模型性能,根据表2设置了四种场景来模拟不同的隐私保护场景。具体来说,实验将隐私预算在1~3之间设置为高级隐私保护,在4~6之间为中等隐私保护,最后大于7的隐私预算为低级隐私保护。将隐私预算阈值ε_{max}设为5,其中隐私预算是根据参与用户的个性化隐私偏好,通过公式7生成的。

case1:参与用户的隐私偏好分布均匀,即设定客户端拥有相同的隐私预算,且隐私预算较小。即,$\varepsilon_1 = \varepsilon_2 = ... = \varepsilon_k <= \varepsilon_{max}$;

case2:参与用户的隐私偏好分布同样均匀,但隐私预算较大。即,$\varepsilon_1 = \varepsilon_2 = ... = \varepsilon_k >= \varepsilon_{max}$;

case3:参与用户的隐私偏好分布符合正态分布,参与用户的预算水平差异较大。即,$\varepsilon \sim N(\mu, \sigma^2)$;

case4:参与用户的隐私偏好分布符合卡方分布,较符合现实情况,即,$\varepsilon \sim x^2(k)$。

在第一种场景中,假设客户端的隐私预算是均匀分配的,且隐私预算小于隐私预算阈值,因此中央聚合服务器没有添加任何噪声,这实际上是本地化差分隐私对联邦学习的保护。因为该场景下参与用户在本地添加噪声,噪声的尺度是大于全局隐私预算阈值的,所以本地的隐私预算直接决定了全局添加的噪声大小。由图2(a)场景1下的分类准确度可以看出,在MNIST数据集上,PDP-FL算法的分类准确率会随着训练轮次和隐私预算ε的增加而增加。由图3(a)场景1下的分类准确度可以看出,在CIFAR-10数据集上,PDP-FL算法的分类准确率也训练轮次的增加而增加,与MNIST数据集不同的是,CIFAR-10数据集由于数据量较大,所以收敛性能较差。第一种情况下,算法认为参与用户的数据已经在本地得到了较好的保护,在中央聚合服务器继续添加噪声会影响模型的训练。因此算法只是在本地添加噪声,此时PDP-FL算法符合(ε_1,δ)-全局差分隐私。

在第二种场景中,假设参与用户的隐私预算是平均分配的,且数值大于隐私预算阈值。这时算法判断参与用户在本地添加的噪声不足,因此需要中央聚合服务器继续添加噪声,以满足全局差分隐私阈值。由图2(b)场景2下的分类准确度可以看出,在MNIST数据集上,PDP-FL算法在多种隐私预算的表现下,分类成功率表现差异不大。这是因为此时的PDP-FL算法符合统一的全局隐私预算阈值。由图3(b)场景2下的分类准确度可以看出,在CIFAR-10数据集上,算法的分类成功率表现差异同样不明显。在第二种场景下,中央服务器需

要向聚合后的模型的权值添加噪声，使其符合$(\varepsilon_{max},\delta)$全局差分隐私。

正态分布是代表连续型随机变量的分布，有两个参数，分别代表平均值的期望μ和描述离散程度的方差σ。因为隐私预算必须是正数，为了满足该条件，进行测试后选择2作为正态分布的方差。因为PDP-FL算法是对中心和本地同时进行保护，为了使得实验更加客观，选择5作为正态分布的均值，隐私偏好通过正态分布函数进行生成，根据公式7生成隐私预算。由图2(c)场景3下的分类准确度可以看出，在第三种场景中，在MNIST数据集上，PDP-FL算法分类的分类准确度在较前面的轮次与无噪声下的分类准确度差距较大，在较后轮次时，添加噪声后的分类准确度与无噪声下的分类准确度差距较小。这是因为联邦学习快速收敛时，对噪声较为敏感。同样，从图3(c)场景3下的分类准确度可以看出，在CIFAR-10数据集上，可以发现，PDP-FL算法分类的分类准确度在模型收敛后，对模型的影响较为有限。

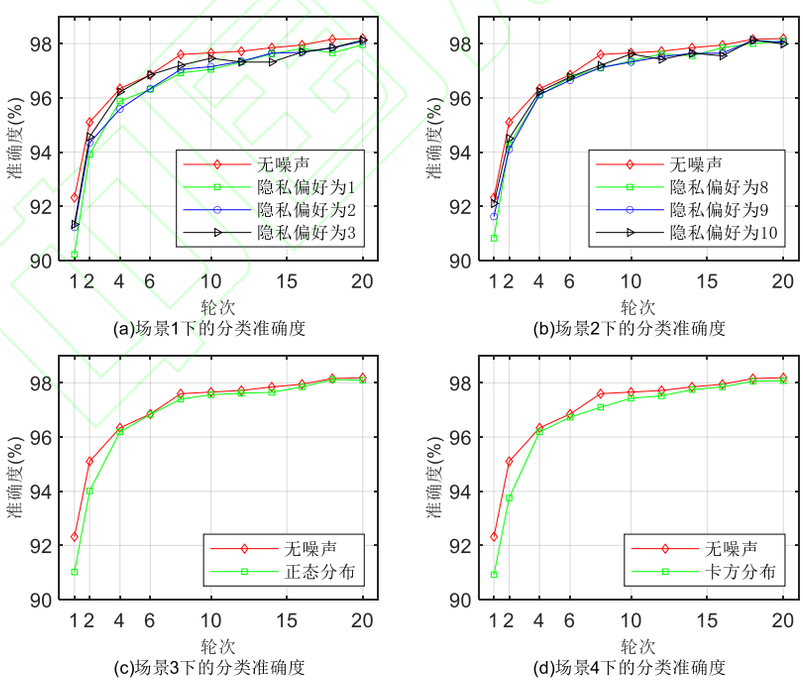
卡方分布是由正态分布构造而成的一种新的分布，在自由度k较大时近似为正态分布，且卡方分布的数值范围符合隐私预算对于正值的要求。为了和正态分布相区别，取自由度为2，此时卡方分布密度函数会较为偏斜，符合现实中某些地域性等的采样过程，隐私偏好通过卡方分布函数进行生成，根据公式7生成隐私预算。由图2(d)场景4下的分类准确度可以看出，在第三种场景中，在MNIST数据集上，PDP-FL算法分类的分类准确度随着轮数的增加而增加。由图3(d)场景4下的分类准确度可以看出，在CIFAR-10数据集上，隐私预算的偏斜对模型的分类准确度造成了较差的影响。在第三种场景中，参与用户的隐私预算分布较为偏斜，大部分参与用户的隐私预算较低，隐私保护需求较高。

4.3 实验参数的影响

本节实验PDP-FL算法的四个参数对分类准确度的影响，即客户端数量、每轮的客户端选择比例、学习率和隐私预算阈值。为了测量每一个参数对分类准确度的影响，采用控制变量法，在改变一个参数的同时，保持其他参数不变。实验采用的情景是符合现实情况的正态分布的隐私预算分布。因为正态分布产生的随机隐私预算可能出现负值的情况，在实验时将方差尽量调小，多次实验后2是比较好的参数。为了尽可能达到两阶段噪声添加测试的目的，设定正态分布的期望为5。

客户端数量N：设置学习速率μ为0.03，每轮的客户端选择比例k的值为0.1，隐私预算阈值ε_{max}为50。如图4(a)客户端数量对分类准确度的影响所示，PDP-FL算法的分类成功率随着客户端数量的增加而增加，这是因为更多的客户端不仅提供了更大的全局数据集进行训练，还降低了由于参数聚合而产生的噪声。可以看出，在模型训练的前几轮，客户端数量的变化会对模型分类准确度产生较大的影响，随着训练轮次的增加，模型训练成功率趋于平稳，这是符合联邦学习训练规律的。

学习率μ：将客户端数量设为100个，每轮的客户端选择比例k的值为0.1，隐私预算阈值ε_{max}设为50。如图4(b)学习率对分类准确度的影响所示，在学习速率为0.03时，PDP-FL算法的分类成功率达到了最佳效果。经过多次实验，学习速率在0.01到0.03之间都可以达到较高的效果，但超过0.05时，模型的分类成功率会显著下降，因此需要根据模型和数据的实际情况对学习率进行判断权衡。学习率在相对较高的水平有利于联邦学习训练模型在训练的前几轮快速收敛，但是过高的学习率会导致模型训练成功率大幅度下降，实际训练过程中需要进行多次实验找出一个比较合适的学习率。



(a)场景1下的分类准确度　(b)场景2下的分类准确度

(c)场景3下的分类准确度　(d)场景4下的分类准确度

图2　多场景下不同隐私预算的分类准确度比较(MNIST)

Fig. 2　Comparison of classification accuracy of different privacy budgets in multiple scenarios(MNIST)

每轮的客户端选择比例k：设定客户端数量为100个，学习速率μ为0.03，隐私预算阈值ε_{max}为50。如图4(c)每轮客户端数量对分类准确度的影响所示，当每轮的客户端选择比例k的值为0.1时，PDP-FL算法的分类成功率达到最佳效果，这是因为对于不同的保护级别，存在一个提高收敛性能的最优k值。这本质上是权衡隐私保护强度的增强对模型参数的影响和在每一轮模型更新中涉及更大的全局训练数据集之间的权衡，同样，这也需要根据具体训练的模型和数据集进行多次测试。

隐私预算阈值ε_{max}：将学习速率设为0.03，每轮的客户端选择比例k的值为0.1，客户端数量设为100。当隐私预算阈值增加时，PDP-FL算法满足的全局差分隐私预算在变大，此时相应的模型的精度会提高。如图4(d)隐私预算阈值对分类准确度的影响所示，当隐私预

算阈值提高时,PDP-FL算法的分类成功率随着隐私预算阈值的增加而增加。

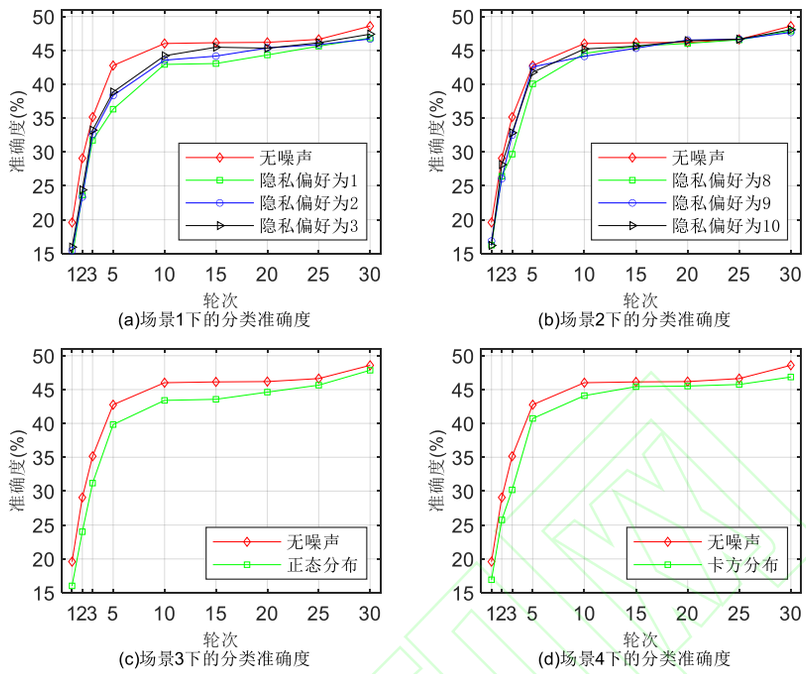


图3 多场景下不同隐私预算的分类准确度比较(CIFAR-10)

Fig. 3 Comparison of classification accuracy of different privacy budgets in multiple scenarios(CIFAR-10)

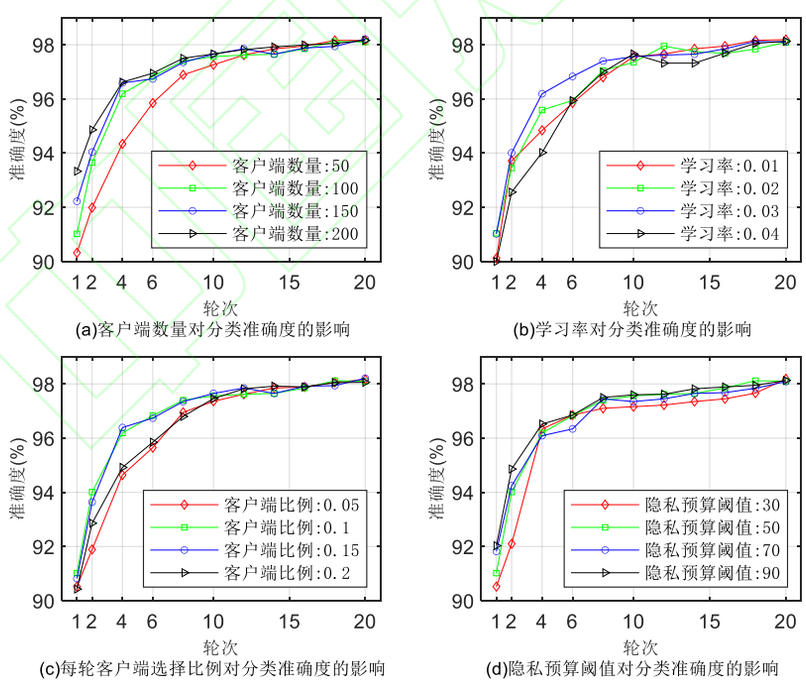


图4 算法参数对分类准确度的影响

Fig. 4 Influence of algorithm parameters on classification accuracy

4.4 与现有方法的比较

如表2和表3所示,将PDP-FL与联邦平均算法(Federated Averaging, FedAvg)[18]、基于差分隐私的联邦学习(Federated Learning with Differential Private, GDP-FL)[26]、基于本地化差分隐私的联邦学习(Federated Learning with Local Differential Private, LDP-Fed)[27]的方法进行了比较。在MNIST数据集上,统一设定客户端的数量为100,聚合次数为2次,每轮的客户端选择比例k的值为0.1,总的隐私预算为50。其中,虽然FedAvg提供了最好的分类准确度是95.1%,但FedAvg不提供隐私保护的效果,攻击者通过获取神经网络的权值,可以获取用户的信息。GDP-FL也具有良好的准确性,但缺点是需要一个可信的中央聚合服务器。一般来说,在个性化隐私保护的场景下,可信的中央聚合服务器是很难获得的。LDP-Fed的分

类准确度为93.7%，LDP-Fed需要设定统一的隐私预算，在本地客户端对数据进行加噪，这大大降低了高预算客户提交模型参数的精度。而PDP-FL的最佳分类准确度（94.5%）接近FedAvg，超过GDP-FL和LDP-Fed。比较表明，PDP-FL算法比现有的基于差分隐私的联邦学习算法具有更好的性能。

在CIFAR-10数据集上，统一设定客户端的数量为100，聚合次数为10次，每轮客户端选择的比例为0.1，总的隐私预算为50。其中，作为不添加噪声FedAvg方法，在CIFAR-10数据集上依然提供了最高的准确度（46.01%），但是FedAvg方法没有提供隐私保护效果。GDP-FL也具有较高的准确度，但GDP-FL采用的中心化差分隐私需要一个可信的中央聚合服务器。相比于LDP-Fed算法，提出的PDP-FL算法具有较高的准确度（45.20%），最接近FedAvg算法的分类准确度。在CIFAR-10数据集上，LDP-FL算法也具有良好的性能。

表2 DL FedAvg，GDP-FL，LDP-Fed和PDP-FL的比较（MNIST）

Tab. 2 COMPARISON BETWEEN GDP-FL，LDP-Fed，PDP-FL and DL FedAvg（MNIST）

隐私保护	算法	准确度/%	损失	场景
无	DL FedAvg	95.1	1.2	无
GDP	GDP-FL	94.2	1.5	无
LDP	LDP-Fed	93.7	1.4	无
		93.9	1.3	Case1
PDP	PDP-FL	94.5	1.2	Case2
		93.8	1.3	Case3
		94.0	1.3	Case4

表3 DL FedAvg，GDP-FL，LDP-Fed和PDP-FL的比较（CIFAR-10）

Tab. 3 COMPARISON BETWEEN GDP-FL，LDP-Fed，PDP-FL and DL FedAvg（CIFAR-10）

隐私保护	算法	准确度/%	损失	场景
无	DL FedAvg	46.01	1.2	无
GDP	GDP-FL	44.22	1.5	无
LDP	LDP-Fed	43.60	1.4	无
		43.55	1.4	Case1
PDP	PDP-FL	45.20	1.2	Case2
		43.40	1.4	Case3
		44.10	1.3	Case4

5 结语

提出了一种两阶段的基于个性化差分隐私的联邦学习算法，在满足个性化隐私保护的前提下，提供了对联邦学习本地和中心的同时保护，并量化了全局的隐私保护强度。首先，解释了基于差分隐私的联邦学习所面临的挑战，以及统一的隐私预算分配不能满足参与用户的需求，浪费了参与用户大部分的隐私预算。然后，详细描述了提出的基于PDP-FL算法，并分析了算法在威胁模型下的鲁棒性。最后，在MNIST和CIFAR-10数据集上，通过设计多个情景下的实验、多个参数对实验的影响和与其他实验的对比，结果表明PDP-FL算法在隐私保护水平和分类准确度上具有良好的表现。

目前PDP-FL算法可以在个性化隐私保护的场景下保证$(\varepsilon_{max},\delta)$的全局差分隐私。但为了保证模型训练的正常进行，采用简单的隐私分级策略来控制用户噪声的最大尺度的方式不够灵活。未来研究工作的方向，计划改进隐私分级的方法，通过结合博弈论等知识更好地建立隐私分级算法。

参考文献（References）

[1] LOPEZ K L, GAGNE C, GARDNER M A. Demand-side management using deep learning for smart charging of electric vehicles [J]. IEEE Transactions on Smart Grid, 2018, 10(3): 2683-2691.

[2] LIN W Y, Hu Y H, TSAI C F. Machine learning in financial crisis prediction: a survey [J]. IEEE Transactions on Systems, Man, and Cybernetics, 2011, 42(4): 421-436.

[3] CHEN S, YU D, ZOU Y, et al. Decentralized wireless federated learning with differential privacy [J/OL]. [2022-3-5]. https://ieeexplore.ieee.org/abstract/document/9693141.

[4] BARRENO M, NELSON B, JOSEPH A D, et al. The security of machine learning [J]. Machine Learning, 2010, 81(2): 121-148.

[5] 张梅舒，徐雅斌. 多维数值型敏感属性数据的个性化隐私保护方法 [J]. 计算机应用，2020，40(02)：491-496.（ZHANG M S, Xu Y B. Personalized privacy protection method for multidimensional numerical sensitive attribute data [J]. Journal of Computer Applications, 2020, 40(02): 491-496.）

[6] 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(03)：1057-1092.（LIU Y X, CHEN H, LIU Y H, et al. Privacy protection technology in federal learning [J]. Journal of Software, 2022, 33(03): 1057-1092.）

[7] WU N, FAROKHI F, SMITH D, et al. The value of collaboration in convex machine learning with differential privacy [C]// Proceedings of the 2020 IEEE Symposium on Security and Privacy. Piscataway: IEEE, 2020: 304-317.

[8] WANG B, YAO Y, SHAN S, et al. Neural cleanse: Identifying and mitigating backdoor attacks in neural networks [C]// Proceedings of the 2019 IEEE Symposium on Security and Privacy. Piscataway: IEEE, 2019: 707-723.

[9] YUAN X, HE P, ZHU Q, et al. Adversarial examples: attacks and defenses for deep learning [J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(9): 2805-2824.

[10] YUAN J, YU S. Privacy preserving back-propagation neural network learning made practical with cloud computing [J]. IEEE Transactions on Parallel and Distributed Systems, 2013, 25(1): 212-221.

[11] BONAWITZ K, IVANOV V, KREUTER B, et al. Practical secure aggregation for privacy-preserving machine learning [C]// Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2017: 1175-1191.

[12] 邱鑫源，叶泽聪，崔翛龙，等. 联邦学习通信开销研究综述[J]. 计算机应用，2022，42(02)：333-342.（QIU J Y, YE Z C, CUI X L, et al. A survey of communication overhead in federated learning [J]. Journal of Computer Applications, 2022, 42(02): 333-342.）

[13] DWORK C, ROTH A. The algorithmic foundations of differential privacy [J]. Found. Trends Theor. Comput. Sci., 2014, 9(3-4): 211-407.

[14] ZHANG J, ZHAO Y, WANG J, et al. Fedmec: improving efficiency of differentially private federated learning via mobile edge computing [J]. Mobile Networks and Applications, 2020, 25(6): 2421-2433.

[15] 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, 2020: 61-66.

[16] ZHAO Y, ZHAO J, YANG M, et al. Local differential privacy-based federated learning for internet of things [J]. IEEE Internet of Things Journal, 2020, 8(11): 8836-8853.

[17] WU X, ZHANG Y, SHI M, et al. An adaptive federated learning

scheme with differential privacy preserving [J]. Future Generation Computer Systems, 2022, 127: 362-372.

[18] 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.

[19] SONG J, WANG W, GADEKALLU T R, et al. Eppda: an efficient privacy-preserving data aggregation federated learning scheme [J/OL]. [2022-3-5]. https://ieeexplore. ieee. org/abstract/document/9721557.

[20] GONG M, FENG J, XIE Y. Privacy-enhanced multi-party deep learning [J]. Neural Networks, 2020, 121: 484-496.

[21] MCMAHAN B, MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data [C]// Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. Cambridge MA: JMLR, 2017: 1273-1282.

[22] 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.

[23] WANG S, HUANG L, NIE Y, et al. Local differential private data aggregation for discrete distribution estimation [J]. IEEE Transactions on Parallel and Distributed Systems, 2019, 30(9): 2046-2059.

[24] ARACHCHIGE P C M, BERTOK P, KHALIL I, et al. Local differential privacy for deep learning [J]. IEEE Internet of Things Journal, 2019, 7(7): 5827-5842.

[25] WANG N, XIAO X, YANG Y, et al. Collecting and analyzing multidimensional data with local differential privacy [C]// Proceedings of the 2019 IEEE 35th International Conference on Data Engineering. Piscataway: IEEE, 2019: 638-649.

[26] GEYER R C, KLEIN T, NABI M. Differentially private federated learning: A client level perspective [EB/OL]. (2017-12-20) [2022-3-5]. https://arxiv. org/pdf/1712. 07557. pdf

[27] 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, 2020: 61-66.

[28] 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.

[29] WU S, YU M, AHMED M A M, et al. Fl-mac-rdp: federated learning over multiple access channels with rényi differential privacy [J]. International Journal of Theoretical Physics, 2021, 60(7): 2668-2682.

[30] LIU R, CAO Y, YOSHIKAWA M, et al. Fedsel: federated sgd under local differential privacy with top-k dimension selection [C]// Proceedings of the International Conference on Database Systems for Advanced Applications. Cham: Springer, 2020: 485-501.

[31] 莫慧凌,郑海峰,高敏,等. 基于联邦学习的多源异构数据融合算法[J]. 计算机研究与发展,2022,59(02):478-487. (MO H L, ZHEN H F, GAO M, et al. Multi-source heterogeneous data fusion algorithm based on Federated learning [J]. Journal of Computer Research and Development, 2022, 59(02): 478-487.)

[32] ZHAO L, WANG Q, ZOU Q, et al. Privacy-preserving collaborative deep learning with unreliable participants [J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 1486-1500.

[33] GIRGIS A, DATA D, DIGGAVI S, et al. Shuffled model of differential privacy in federated learning [C]// Proceedings of the 24th International Conference on Artificial Intelligence and Statistics. Cambridge MA: JMLR, 2021: 2521-2529.

[34] ZHANG L, ZHU T, XIONG P, et al. A robust game-theoretical federated learning framework with joint differential privacy [J/OL]. [2022-3-5]. https://ieeexplore. ieee. org/abstract/document/9669031.

[35] AVENT B, KOROLOVA A, ZEBER D, et al. Blender: enabling local search with a hybrid differential privacy model [C]// Proceedings of the 26th USENIX Security Symposium. Vancouver , BC, Canada: USENIX Association, 2017: 747-764.

[36] HU R, GUO Y, LI H, et al. Personalized federated learning with differential privacy [J]. IEEE Internet of Things Journal, 2020, 7(10): 9530-9539.

[37] YANG G, WANG S, WANG H. Federated learning with personalized local differential privacy [C]// Proceedings of the 2021 IEEE 6th International Conference on Computer and Communication Systems. Piscataway: IEEE, 2021: 484-489.

[38] LI H, XIONG L, JI Z, et al. Partitioning-based mechanisms under personalized differential privacy [C]// Proceedings of the Pacific-asia Conference on Knowledge Discovery and Data Mining. Cham: Springer, 2017: 615-627.

[39] INAN A, GURSOY M E, SAYGIN Y. Sensitivity analysis for non-interactive differential privacy: bounds and efficient algorithms [J]. IEEE Transactions on Dependable and Secure Computing, 2017, 17(1): 194-207.

[40] JORGENSEN Z, Yu T, CORMODE G. Conservative or liberal? Personalized differential privacy [C]// Proceedings of the 2015 IEEE 31st International Conference on Data Engineering. Piscataway: IEEE, 2015: 1023-1034.

[41] ZHANG Y, QU Y, GAO L, et al. GPDP: Game-Enhanced Personalized Differentially Private Smart Community [C]// Proceedings of the 2021 IEEE International Conferences on Internet of Things and IEEE Green Computing and Communications and IEEE Cyber, Physical and Social Computing and IEEE Smart Data and IEEE Congress on Cybermatics. Piscataway: IEEE, 2021: 238-243.

[42] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[43] KRIZHEVSKY A, HINTON G. Learning multiple layers of features from tiny images [J/OL]. [2022-3-5]. https://www. cs. toronto. edu /~kriz/learning-features-2009-TR. pdf.

YIN Chunyong, born in 1977, Ph. D., professor, Ph. D. supervisor. His research interests include cyberspace security, big data mining, privacy protection, artificial intelligence, new computing.

QU Rui, born in 1999, M. S. candidate. His research interests include differential privacy, federated learning.