2025 年 第 25 卷 第 7 期
2025, 25(7):02849-07
科 学 技 术 与 工 程
Science Technology and Engineering
ISSN 1671—1815
CN 11—4688/T

DOI:10.12404/j.issn.1671-1815.2308502
引用格式:赵婵婵, 马坤明, 石宝, 等. 基于差分隐私的自适应联邦学习隐私保护方案[J]. 科学技术与工程, 2025, 25(7): 2849-2855.
Zhao Chanchan, Ma Kunming, Shi Bao, et al. Adaptive privacy protection scheme for federated learning based on differential privacy[J]. Science Technology and Engineering, 2025, 25(7): 2849-2855.

基于差分隐私的自适应联邦学习隐私保护方案

赵婵婵, 马坤明, 石宝*, 杨星辰, 李燕
(内蒙古工业大学信息工程学院, 呼和浩特 010080)

摘 要 随着对联邦学习的深入研究,发现联邦学习中的隐私保护策略并不能完全保护用户的隐私安全,并且在联邦学习训练过程中存在模型收敛困难的问题。针对以上问题,提出了一种自适应差分隐私机制(adaptive differential privacy, DP-AdaMod)。首先,利用自适应学习率算法调整模型训练过程,避免模型出现波动和过拟合现象,从而提高模型训练的效率和性能。其次,引入差分隐私技术,通过对模型梯度添加噪声来确保联邦学习的隐私安全。同时,使用 Moment Accountant 机制进行隐私损失的精确计算,有助于平衡隐私保护性能和精度,从而进一步增强了系统的安全性。最后,通过仿真实验验证所提方案的有效性。结果表明该方案在准确率、隐私预算消耗等方面展现出较优性能。

关键词 联邦学习; 差分隐私; 隐私保护; 自适应
中图法分类号 TP309; **文献标志码** A

Adaptive Privacy Protection Scheme for Federated Learning Based on Differential Privacy

ZHAO Chan-chan, MA Kun-ming, SHI Bao*, YANG Xing-chen, LI Yan
(College of Information Science and Engineering, Inner Mongolia University of Technology, Hohhot 010080, China)

[**Abstract**] With the deepening research on federated learning, it has been observed that the privacy protection strategies employed within federated learning fall short of fully guaranteeing the security and confidentiality of user data. Moreover, the training process in federated learning encounters challenges regarding model convergence. In response to these aforementioned issues, an innovative solution termed adaptive differential privacy (DP-AdaMod) was proposed. Primarily, the model training process was fine-tuned by incorporating an adaptive learning rate algorithm to mitigate model fluctuations and the adverse effects of overfitting. Consequently, this enhancement led to improved training efficiency and optimal performance. Secondly, the application of differential privacy techniques ensured the privacy security in federated learning through the deliberate introduction of noise into the model gradients. Additionally, accurate quantification of privacy loss was achieved by implementing the moment accountant mechanism, facilitating a balanced trade-off between privacy preservation and analytical accuracy. This meticulous approach served to fortify system security. Lastly, the efficacy of the proposed solution was ascertained through comprehensive simulation experiments. The results substantiate the superior performance of the proposed method, evident by its exceptional accuracy, efficient utilization of privacy budget, and other notable facets.

[**Keywords**] federated learning; differential privacy; privacy protection; adaptive

新兴的技术如云计算、边缘计算和物联网的发展导致全球数据量急剧增长,这给数据的流通和共享带来了巨大的挑战,特别是数据安全方面的威胁变得日益严峻。

传统中心化机器学习通过将数据集中上传至中央服务器进行模型训练,这会导致数据传输过程容易被攻击者截获,造成隐私泄露问题。McMahan 等[1]在 2016 年首次提出联邦学习(federated learning, FL)的概念,并于 2017 年提出联邦平均算法(federal average algorithm, FedAvg)。联邦学习是一种分布式机器学习方法,旨在实现在不将参与方的私有数据上传到中央服务器的情况下,以隐私保护的方式交换中间模型参数,从而协作完成分布式任务。当前,联邦学习的研究主要集中在如何保护隐私和确保安全性方面,而差分隐私技术是一个重要的隐私保护手段,其采用对数据添加噪声的方法来

收稿日期: 2023-10-31 **修订日期**:2024-07-26
基金项目: 内蒙古自治区高等学校科学研究项目(NJZY22382);内蒙古自治区直属高校基本科研业务费项目(JY20240010, JY20230082);内蒙古自治区自然科学基金(2023LHMS06016);内蒙古工业大学学科学研究项目(BS201936)
第一作者: 赵婵婵(1982—),女,汉族,山西临汾人,博士,副教授。研究方向:隐私保护、边缘计算。E-mail:cczhao@imut.edu.cn。
* **通信作者**: 石宝(1982—),男,蒙古族,内蒙古兴安盟人,博士,教授。研究方向:人工智能。E-mail:kshibao@163.com。

保护隐私。自从 2016 年 Dwork[2] 首次提出差分隐私技术(differential privacy,DP)的概念以来,研究者们根据这一概念进行了大量的研究,提出了许多满足差分隐私要求的算法。Wei 等[3] 为了增强联邦学习的隐私保护性能,提出了一种本地化差分隐私机制,该方法通过在每个联邦学习参与方本地添加噪声实现隐私保护,并分析了在相同隐私预算消耗的情况下,参与方数量与全局模型收敛之间的关系。Rü 等[4] 在研究中将联邦学习与差分隐私技术相结合,提出了一种差分隐私模糊交替方向乘子算法 DP-IADMM(differential privacy inexact alternating direction method of multipliers),该算法解决了随机噪声对目标扰动的子问题。Huang 等[5] 针对联邦学习数据分布不平衡的问题,提出了一种差分隐私联邦学习框架(differential privacy federated learning,DP-FL),通过使用新的模型聚合算法,引入权重调整机制以及应用差分隐私机制保护隐私,从而确保在不平衡数据分布情况下的有效性。Hu 等[6] 提出了一种基于差分隐私技术的轻量级联邦学习模型聚合方法 DP-PASGD(differential privacy periodic averaging SGD),通过引入差分隐私技术,提升了隐私保护性能,并通过优化参数聚合过程,降低了通信和计算开销。Zhang 等[7] 提出了一种基于移动边缘计算的联邦学习框架(federated learning mobile edge computing,FedMEC),通过将部分计算任务迁移到边缘服务器上,并引入差分隐私技术,提高了差分隐私联邦学习的效率和隐私保护性能。Leng 等[8] 使用了许多优化算法,改善由于联邦学习模型固定的学习率导致收敛过程出现局部收敛的情况。江欣俞等[9] 提出了一种基于图神经网络的兴趣点推荐的隐私保护框架,并采用可变动态梯度的客户端差分隐私算法,达到了边优化边反馈的效果。李洋等[10] 提出一种可实现双向自适应差分隐私的联邦学习方案(federated learning mobile bidirectional adaptive differential privacy,FedBADP)可以对数据之间传输的梯度进行自适应加噪,但在同步联邦学习训练方式下,仍存在参数泄露的风险。Xu 等[11] 通过使用自适应梯度下降来加速模型收敛和降低隐私损失,但该方案的模型准确率较低。

综上所述,当前差分隐私与联邦学习的结合,虽然对联邦学习的隐私保护能力有了很大提升,但是由于噪声会对模型的精度造成影响,很难在模型的隐私性和准确性达到一个平衡。为此,提出一种自适应差分隐私优化算法,可以自适应调整学习率,从而有效降低差分隐私添加噪声对模型精度的影响。本文的重要工作如下。

(1)提出了一种自适应差分隐私优化算法。它可以根据历史梯度信息自适应地调整学习率,以更好地适应不同参数的变化。这使得在联邦学习中,模型能够更好地适应不同客户端的数据特点,提高全局模型的准确性。

(2)引入差分隐私保护机制,确保在联邦学习过程中隐私得到保护。通过应用差分隐私技术,对模型更新的梯度添加噪声,从而保护具体客户端的隐私。这使得本文方案可以在联邦学习中更好地平衡模型精度和隐私保护需求。

1 相关知识

1.1 联邦学习

联邦学习(FL)是一种分布式机器学习方法,可以根据参与方数据集的特征分为以下 3 类:横向联邦学习、纵向联邦学习和联邦迁移学习[12-13]。

图 1 为典型联邦学习系统框架,由多个参与方和一个聚合服务器组成。每个参与方都持有数据特征完整的且数据样本之间几乎没有交集的本地数据集。通过聚合服务器的协调,参与方可以联合起来训练一个更加高效的全局模型。具体训练过程如下。

(1)模型初始化:联邦学习开始之前,需要初始化一个全局模型。聚合服务器初始化全局模型并选择适合于训练的客户端。

(2)数据分发:服务器将初始化的全局模型发送给参与方用作本地训练的初始模型。

(3)本地训练:参与方接收到全局模型后,使用本地数据集训练本地模型。

(4)模型聚合:每个参与方完成本地训练后,聚合服务器收集各参与方的模型参数并使用聚合算法进行模型聚合。

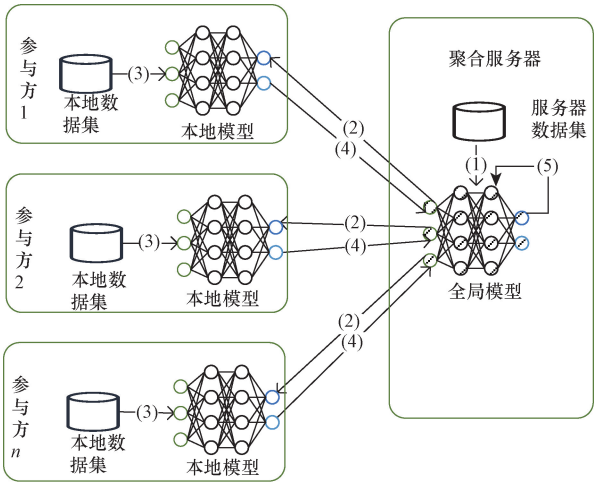

图 1 联邦学习框架

Fig. 1 Federated learning framework

(5)模型更新:聚合服务器将聚合后的全局模型发送给所有参与方,用于更新本地模型。

(6)重复迭代:重复步骤(2)~(5),直至全局模型收敛或者达到预定停止条件。

1.2　差分隐私

差分隐私是一种隐私保护技术,旨在通过在原始数据集或数据交换过程中添加精心设计的噪声,以保护数据隐私。差分隐私可以分为中心化差分隐私和本地化差分隐私[2]。中心化差分隐私是指在集中式环境中,通过在中央服务器对数据添加噪声以保护数据隐私。本地化差分隐私则是指在分布式环境中,将数据处理过程全部在本地设备完成,用户可以在本地对数据进行加噪处理以保护敏感信息。本文中主要应用本地化差分隐私思想,本地化差分隐私的实现如图2所示。

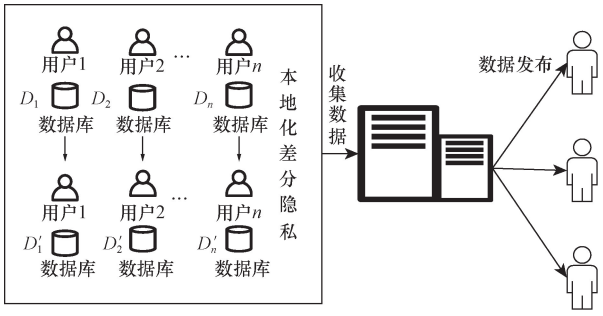

图2　本地化差分隐私

Fig. 2　Localization of differential privacy

定义1[14] (ε,δ)-差分隐私。对于给定 $\varepsilon>0$ 和 $\delta>0$,若其算法 M 对于任意相邻数据集 D 和 D',其输出结果 S 满足关系式

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

则证明 M 满足 (ε,δ)-差分隐私。其中,ε 为隐私预算,δ 是一个常数项,用来衡量隐私保护机制中允许的随机性额外噪声的影响。

定义2[14]　隐私敏感度。对于任意查询函数 f,将数据集 D 映射到输出空间中,隐私敏感度可以表示为

$$\Delta f=\max_{D,D'}\|f(D)-f(D')\|_p \tag{2}$$

式(2)中:$\|\cdot\|_p$ 表示 L_p 范数,即对于 D 和 D',函数 f 的最大变化值。隐私敏感度是差分隐私中的关键定义,用来衡量或者量化查询函数 f 在 D 和 D' 上输出结果变化的敏感程度。隐私敏感度越高,则隐私泄露的风险也会越大。为了降低隐私泄露风险,添加的噪声也会相应变大。

定义3[14]　高斯噪声。对于查询函数 f,向其输出 $f(D)$ 中添加高斯噪声,实现 (ε,δ)—— 差分隐私。即

$$M(D)=f(D)+N[0,(\Delta f\sigma)^2\boldsymbol{I}] \tag{3}$$

式(3)中:高斯分布的均值为0;协方差为 $(\Delta f\sigma)^2\boldsymbol{I}$,$\boldsymbol{I}$ 为单位矩阵。

2　方案设计

2.1　自适应差分隐私优化算法(DP-AdaMod)

学习率和迭代次数是机器学习过程之前需要设置的超参数。传统随机梯度下降算法(stochastic gradient descent,SGD)[15-16]由于固定的学习率,往往会在训练过程中减慢收敛速度并且有很强的震荡性。为了更好地适应不同维度目标函数的优化过程,自适应学习率的梯度下降算法开始被研究者们研究和使用。Adam[17-18]兼具了AdaGrad[19]和RMSProp[20]的优点,它也可以利用一阶矩均值计算自适应学习率;不同的是,Adam还充分利用了梯度的二阶矩均值。然而,尽管Adam算法具备这些优点,但是Adam仍然存在着在某些情况下不收敛和波动大的情况。AdaMod优化算法[21]是基于Adam算法的思想进行改进的,AdaMod算法作为一种自适应学习率的算法,能够在训练开始时就控制自适应学习率的方差,以确保训练初期的稳定性。因此,在AdaMod优化算法思想基础上进行改进,提出了一种自适应差分隐私优化算法(DP-AdaMod),通过在梯度中加入噪声保护隐私,并利用自适应学习率加快模型收敛,有效平衡了模型的隐私性和准确性。

AdaMod优化算法计算一阶矩的步骤如下:

(1)在每轮迭代过程中,算法会从训练集中随机选出 m 个样本,然后通过计算先前一阶矩的指数移动平均值计算当前梯度的一阶矩。即

$$m_t=\beta_1 m_{t-1}+(1-\beta_1)g_t \tag{4}$$

式(4)中:m_t 为第 t 次迭代时梯度的一阶矩估计值;β_1 为衰减率,其值基于0和1;g_t 为梯度。

(2)AdaMod算法通过使用梯度 g_t 平方的指数移动平均数来估算梯度的二阶矩。具体来说,在每次迭代过程中,AdaMod优化算法通过计算先前二阶矩的指数移动平均值来估计当前梯度的二阶矩。该估算值将用于调整学习率,以适应梯度变化。初始时,二阶矩估计为零向量。然后计算进行二阶矩,公式为

$$v_t=\beta_2 v_{t-1}+(1-\beta_2)g_t^2 \tag{5}$$

式(5)中:β_2 为衰减率;v_t 为第 t 次迭代时梯度平方的二阶矩估计值。通过计算一阶矩和二阶矩,AdaMod算法能够自适应地调整梯度更新步长,并在训练中动态地平衡参数更新的速度和方向,以提高训练效果。

(3)AdaMod算法通过引入修正项以减少初始迭代时估计偏差的影响。具体而言,在每次迭代中,将 β_1^t 和 β_2^t 分别用于一阶矩和二阶矩的估算,得到修正后的一阶矩 $\hat{m}_t$ 和二阶矩 $\hat{v}_t$ 为

$$\begin{cases}\hat{m}_t = \dfrac{m_t}{1-\beta_1^t} \\ \hat{v}_t = \dfrac{v_t}{1-\beta_2^t}\end{cases} \tag{6}$$

通过引入修正量,AdaMod 算法能够减少初始迭代时估计偏差的影响,从而更准确地估计一阶矩和二阶矩,提高算法在初始阶段的稳定性和性能。

(4)最后,在每一轮的迭代中,参数值 θ_t 将通过学习率 η_t 乘 $\hat{m}_t$ 来计算,即

$$\theta_{t+1} \leftarrow \theta_t - \eta_t \hat{m}_t \tag{7}$$

AdaMod 算法通过对学习率进行指数加权平均来计算当前指数的滑动平均值,即

$$s_t = \beta_3 s_{t-1} + (1-\beta_3)\eta_t \tag{8}$$

$$s_t = (1-\beta_3)[s_{t-1} + \beta_3 s_{t-2} + \cdots + \beta_3^{t-1} s_0] \tag{9}$$

式中:s_t 为指数滑动平均值;β_3 为记忆时长参数,其值小于 1。

根据式(9),可以计算出当前的指数滑动平均值。然后,比较当前指数滑动平均值 s_t 和学习率 η_t 的值,将较小的值作为当前的学习率。这样做有效避免了学习率过高的问题。通过不断重复上述步骤,可以保证学习率一直受到当前指数滑动平均值的限制,使其更稳定地适应模型训练需求。

DP-AdaMod 算法的主要步骤包括梯度剪裁、学习率计算和添加噪声。梯度剪裁是一种常见的模型优化技术,限制梯度范围在适当的范围内,以防止梯度爆炸问题,并加速模型的收敛。在 DP-AdaMod算法中,对每个梯度向量的L_2范数应用梯度剪裁。梯度剪裁的公式为

$$g_t(x_i) \leftarrow \frac{g_t(x_i)}{\max[1, \| g_t(x_i) \|_2 / C]} \tag{10}$$

式(10)中:g_t 为当前梯度;C 为裁剪阈值。

噪声添加的公式为

$$\tilde{g}_t \leftarrow \frac{1}{L}\sum_i [\hat{m}_t + N(0, \sigma_t^2 C^2 \boldsymbol{I})] \tag{11}$$

本文中使用高斯噪声 $N(0,\sigma^2C^2\boldsymbol{I})$,噪声的概率分布是标准差为 $C\sigma$、均值为 0 的高斯分布。对于梯度向量 g_t,如果其L_2范数 $\| g_t \|_2 \leqslant C$,则保留原始梯度g_t;如果其L_2范数 $\| g_t \|_2 > C$ 时,则将梯度裁剪为 C,即 $g_t = C$。因此,可以通过设置裁剪阈值 C 的值调整后续添加的噪声大小,这意味着后续添加的噪声大小与裁剪阈值 C 的大小密切相关。这种方式允许能够对梯度向量进行裁剪,并且能够对噪声大小进行控制,从而实现隐私保护性能和模型精度之间的平衡。

DP-AdaMod 算法计算学习率的过程为:首先,计算一阶矩和二阶矩的估计值。其次,使用一阶矩的估计值和二阶矩的估计值来计算出学习率 η_t。然后,计算当前的指数滑动平均值 s_t。最后对比 η_t 和 s_t 值的大小,将二者中值较小的作为当前学习率。

本文算法的具体实现如下:

算法 1 自适应差分隐私算法(DP-AdaMod 算法)

Input:初始化参数 θ_0,初始化参数样本 $\{x_1, x_2, \cdots, x_N\}$,学习 $\{\eta_t\}_{t=1}^T$,噪声参数 σ,批次大小 L,衰减参数 $\{\beta_1, \beta_2, \beta_3\}$,梯度裁剪阈值 C,迭代次数 T,损失函数 $L(\theta) = \frac{1}{N}\sum_i L(\theta, x_i)$。

Initialize: $m_0 = 0$, $v_0 = 0$, $s_0 = 0$,

for $t = 1$ to T do:

Step1: 计算梯度 $g_t(x_i) \leftarrow \nabla_{\theta_t} L(\theta, x_i)$

Step2:梯度修剪 $g_t(x_i) \leftarrow \dfrac{g_t(x_i)}{\max(1, \| g_t(x_i) \|_2 / C)}$

Step3:计算学习率和梯度累计值: $\bar{g}_t \leftarrow \sum_{i=0}^{m} g_t(x_i) m_t = \beta_1 m_{t-1} + (1-\beta_1)\bar{g}_t$

$v_t = \beta_2 v_{t-1} + (1-\beta_2)\bar{g}_t^2$

$\hat{m}_t = m_t / (1-\beta_1^t)$

$\hat{v}_t = v_t / (1-\beta_2^t)$

$\eta_t = \alpha_t / (\sqrt{\hat{v}_t} + \varepsilon)$

$s_t = \beta_3 s_{t-1} + (1-\beta_3)\eta_t$

$\hat{\eta}_t = \min(\eta_t, s_t)$

Step4:噪声添加 $\tilde{g}_t \leftarrow \frac{1}{L}\sum_i [\hat{m}_t + N(0, \sigma_t^2 C^2 \boldsymbol{I})]$

Step5:参数更新 $\theta_t = \theta_{t-1} - \hat{\eta}_t \tilde{g}_t$

end for

Out Put: θ_t

2.2 全局模型的更新优化

中心服务器并不会收集所有客户端的数据,而是通过收集梯度更新来更新中心模型,此过程可以在一定程度上防止攻击者反推出模型数据。对于每个参与计算的客户端,每轮前都会从服务器获取最新的模型参数作为初始模型参数,并利用本地模型计算梯度更新并发送给中心服务器。中心模型的更新过程主要包括如下几个关键步骤。

(1)中央服务器选择适合于训练的客户端,同时发送初始参数模型并开放下载权限。

(2)参与方使用本地数据集更新参数并计算更新的梯度。在这一步骤中使用自适应梯度下降算法来保证良好收敛性。

(3)避免单个节点过度贡献,对一些参与方更新的梯度进行缩放,使模型可以更好地综合各参与方的特征,从而提高整体模型性能。

(4)将差分隐私噪声添加到单个通信过程中的整体聚合更新中。

(5)将梯度更新的结果发送到中央服务器,同时,将更新的全局模型广播给参与方。图 3 为本文中总体架构。

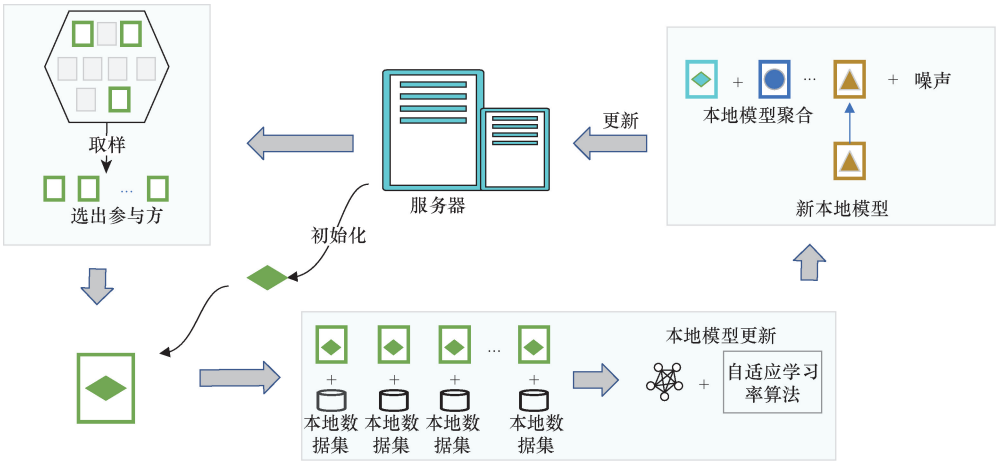

图 3　总体框架

Fig. 3　Overall framework

2.3　隐私性分析

在 DP-AdaMod 算法中,每进行一次梯度计算或者更新一次参数都会产生隐私损失。因此,在一定隐私预算条件下,能够准确计算算法的隐私损失尤为重要。本文中利用 Abadi 等[22]提出的时刻统计方法(moment accountant)来进行隐私统计。该方法可以帮助参与方更好地平衡隐私性和准确率,可以增强系统的安全性。相关定义如下:

定义 4　隐私损失[22]。对于随机算法 M,其隐私损失是通过计算 M 在相邻数据集 D 和 D' 上产生相同输出概率之比的对数来表示,即

$$c(S;M,A,D,D') \triangleq \lg \frac{\Pr[M(A,D)=S]}{\Pr[M(A,D')=S]} \tag{12}$$

式(12)中:A 为辅助变量。

定义 5　时刻统计[22]。算法 M 在 λ 时刻的时刻统计为

$$\alpha(\lambda) \triangleq \max_{D,D'} \lg E_{S\sim M(D)}\{\exp[\lambda c(S,M,D,D')]\} \tag{13}$$

定理 1　组合特性[22]。给定算法 M 和一系列子算法 $M_1,M_2,\cdots,M_n$,其对任意时刻 λ 的时刻统计上界等于所有子算法在该时刻的时刻统计之和。

$$\alpha_M(\lambda) \leqslant \sum_{i=1}^{N} \alpha_{M_i}(\lambda) \tag{14}$$

式(14)中:$\alpha_M(\lambda)$ 为算法 M 在 λ 时刻的时刻统计上届;$\alpha_{M_i}(\lambda)$ 为子算法 M_i 在 λ 时刻的时刻统计上界。

定理 2　差分隐私边界[22]。对于任意 $\varepsilon>0$,若满足

$$\delta = \min_{\lambda} \exp[\alpha_M(\lambda) - \lambda\varepsilon] \tag{15}$$

则算法 M 满足 (ε,δ) -差分隐私。

根据定义 5 计算得到 DP-AdaMod 算法在 λ 时刻统计的值为 $\alpha_M(\lambda;A,D,D') \triangleq \max_{D,D'} \lg E_{S\sim M(A,D)}\{\exp[\lambda c(S;M,A,D,D')]\}$

根据组合特性,假设算法的时刻统计为 $\alpha(\lambda)$,可以推出

$$\alpha(\lambda) \leqslant \sum_{t}^{T} \sum_{i=1}^{N} \alpha_{i,j}(\lambda) \tag{16}$$

式(16)中:$\alpha_{i,j}(\lambda)$ 在梯度上添加高斯噪声 $\xi \sim N(0,\sigma^2C^2\boldsymbol{I})$;$T$ 为总训练轮数;N 为设备总数。T 和 N 与 DP-AdaMod 算法的隐私损失成正比。$\alpha_{i,j}(\lambda)$ 的计算过程如下。

令 μ_0 表示高斯分布 $N(0,\sigma^2C^2\boldsymbol{I})$ 的概率密度函数,μ_1 表示高斯分布 $N(1,\sigma^2C^2\boldsymbol{I})$ 的概率密度函数,μ 表示同时混合 μ_0 和 μ_1 的高斯分布,即 $\mu=(1-q)\mu_0+q\mu_1$,其中 q 为进行本地训练时的抽样概率。根据式(12)和式(13)可以推出 $\alpha_{i,j}(\lambda)$ 的表达式 $\alpha_{i,j}(\lambda)=\lg\max(E_1,E_2)$,其中 E_1 和 E_2 可以分别表示为 $E_1=E_{x\sim\mu_0}\{[\mu_0(x)/\mu(x)]^2\}$,$E_2=E_{x\sim\mu}\{[\mu(x)/\mu_0(x)]^2\}$。根据上述推导可得隐私损失边界为 $\alpha_{i,j}(\lambda) \leqslant q^2\lambda(\lambda+1)/(1-q)\sigma^2+O(q^3/\sigma^3)$。同时,因为本地设备添加的噪声 $\xi\sim N(0,\sigma^2C^2\boldsymbol{I})$ 相同,根据计算得到的 $\alpha(\lambda)$,利用定理 2 推出满足 $(\varepsilon,\delta)=\min_{\lambda}\exp[\alpha(\lambda)-\lambda\varepsilon]$ -差分隐私。

3　实验及结果分析

本文中网络结构采用 LeNet-5 卷积神经网络。由 3 个 5×5 的卷积层(C1、C3、C5),两个 2×2 的池化层(S2、S4)和一个输出层组成,输出层共有 10 个神经元,输出为 0~9。实验采用 MNIST 数据集,该数据集被广泛应用于联邦学习的测试中。MNIST 是一个手写数字识别数据集,包含大约 70 000 张

28×28的灰度图像,标签取值为0~9。该数据集由训练集和测试集两部分组成。其中,训练集包含60 000张训练图像,用于训练模型参数;测试集包含10 000个测试图像,用于评估模型性能。

3.1 准确率评估

为了评估DP-AdaMod算法的准确率,在隐私预算分别为ε = 0.5 ,2,8条件下,将DP-AdaMod算法与DPSGD和DPAdam算法进行准确率的比较。实验设置梯度裁剪阈值为C = 1、噪声参数σ = 4、初始学习率η = 0.001。结果如图4所示。

如图4所示,在隐私预算ε的值分别为0.5、2和8的条件下,提出的DP-AdaMod算法均取得了不错的效果,准确率优于DPSGD和DPAdam算法。实验结果表明,在引入差分隐私技术的联邦学习训练过程中,与固定学习率的梯度下降算法相比,DP-AdaMod算法能够更好地平衡隐私性和准确性。本文算法可以有效防止学习率过大情况,使得每一次学习率自适应更新的值更加精确,保证联邦学习在加噪场景依旧可以保持较好的性能。

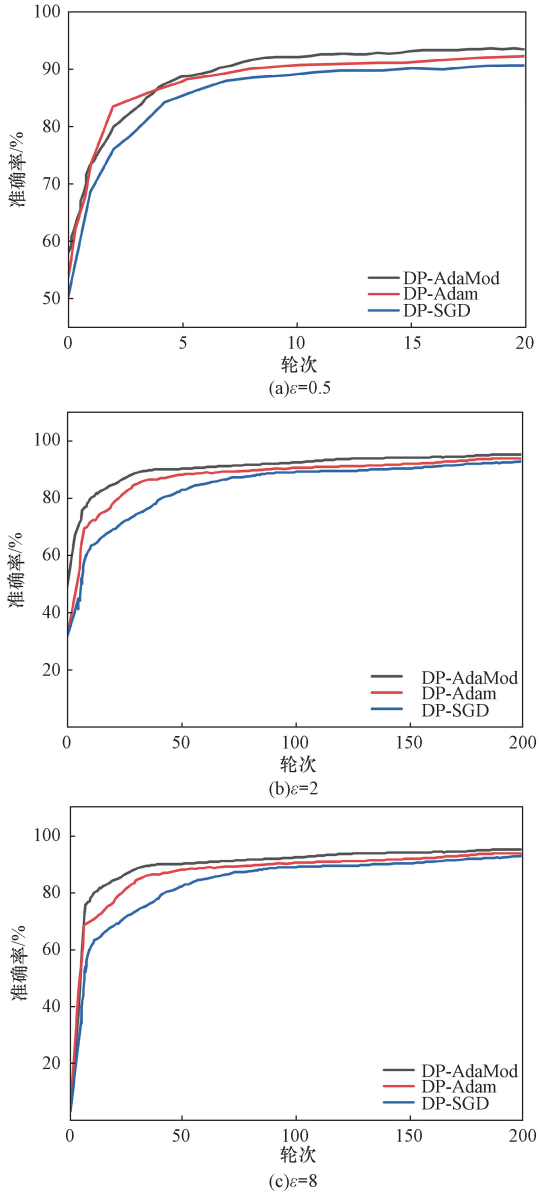

图4 不同隐私预算条件下准确率对比

Fig. 4 Comparison of accuracy under different privacy budget conditions

3.2 隐私预算消耗

为了衡量DP-AdaMod算法在减少隐私预算消耗方面的作用,采用DPSGD和DPAdam作为对比。表2是在准确率分别为88%、90%、92%和94%时上述3种算法的隐私预算消耗。其中ε_1、ε_2、ε_3分别代表DP-SGD、DP-Adam和DP-AdaMod3种算法消耗的隐私预算,减少率1代表DP-AdaMod算法相对于DP-SGD隐私预算消耗减少率,减少率2代表DP-AdaMod算法相对于DP-Adam算法隐私预算消耗减少率。

表2 相同准确率条件下消耗隐私预算对比

Table 2 Comparison of privacy budget consumption under the same accuracy condition

数据集	准确率/%	δ	ε_1	ε_2	ε_3	减少率1/%	减少率2/%
MNIST	88	10^{-5}	0.71	0.62	0.56	21.13	8.94
	90		1.28	1.09	0.92	28.04	15.51
	92		1.78	1.32	1.23	30.90	6.81
	94		5.73	3.68	2.98	47.99	19.02

由表2可以看出,DP-AdaMod算法在MNIST数据集上相比DP-SGD算法平均减少了约32%的隐私预算,比DP-Adam算法平均减少了12.57%的隐私预算。结果表明,DP-AdaMod算法在达到相同准确率的条件下,减少了隐私预算消耗。随着隐私预算的增加,隐私保护程度会降低。因此,在实现相同准确率的前提下,与其他两种算法相比,DP-AdaMod算法的隐私预算消耗较低,能够提供更高的隐私保护程度。

4 结论

针对联邦学习存在的隐私安全问题和收敛困难问题,提出了一种自适应差分隐私机制。经过仿真实验验证,得到如下结论。

(1)提出的方法可以根据不同的训练阶段自适应地调整学习率,提高了模型收敛效率,解决了联邦学习模型收敛困难的问题。

(2)引入了差分隐私技术,通过对训练梯度添加噪声有效保护了联邦学习的隐私安全。

(3)相较于其他先进方法,提出的自适应差分隐私机制可以在消耗较少的隐私预算的情况下,取得不错的准确率。

参 考 文 献

[1] Mcmahan H 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. Fort Lauderdale: JMLR Workshop and Conference Proceedings, 2017: 1273-1282.

[2] Dwork C. Differential privacy[C]//International Colloquium on automata, languages, and Programming. Berlin: Springer Berlin Heidelberg, 2006: 1-12.

[3] Wei K, Li J, Ding M, et al. User-level privacy preserving federated learning: Analysis and performance optimization[J]. IEEE Transactions on Mobile Computing, 2021, 21(9): 3388-3401

[4] Rü M, Kim K. Differentially private federated learning via inexact admm[J]. arxiv preprint arxiv: 2106. 06127, 2021.

[5] 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.

[6] Hu R, Guo Y, Ratazzi E P, et al. Differentially private federated learning for resource-constrained internet of things[J]. arXiv preprint arXiv: 2003. 12705, 2020.

[7] 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: 2421-2433.

[8] Leng J, Ye S, Zhou M, et al. Block chain-secured smart manufacturing in industry 4. 0: a survey[J]. IEEET ransactions on Systems, Man, and Cybernetics: Systems, 2020, 51(1): 237-252.

[9] 江欣俞, 李晓会, 秦若婷, 等. 基于图神经网络的兴趣点推荐的隐私保护框架[J]. 科学技术与工程, 2023, 23(17): 7407-7419.
Jiang Xinyu, Li Xiaohui, Qin Ruoting, et al. Privacy protection framework based on recommendation of points of interest[J]. Science Technology and Engineering, 2023, 23(17): 7407-7419.

[10] 李洋, 徐进, 朱建明, 等. 可实现双向自适应差分隐私的联邦学习方案[J]. 西安电子科技大学学报, 2024, 51(3): 158-169.
Li Yang, Xu Jin, Zhu Jianming, et al. A federated learning scheme that can achieve bidirectional adaptive differential privacy[J]. Journal of Xidian University, 2024, 51(3): 158-169.

[11] Xu Z, Shi S, Liu A X, et al. An adaptive and fast convergent approach to differentially private deep learning[C]//IEEE INFOCOM 2020-IEEE Conference on Computer Communications. New York: IEEE, 2020: 1867-1876.

[12] Ye D, Yu R, Pan M, et al. Federated learning in vehicular edge computing: a selective model aggregateon approach[J]. IEEE Access, 2020, 8: 23920-23935.

[13] Liu Y, Kang Y, Xing C, et al. A secure federated transfer learning framework[J]. IEEE Intelligent Systems, 2020, 35(4): 70-82.

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

[15] Yang Q, Liu Y, Chen T, et al. Federated machine learning: concept and applications[J]. ACM Transactions on Intelligent systems and Technology, 2019, 10(2): 1-19.

[16] Mercier Q, Poirion F, Désidéri J A. A stochastic multiple gradient descent algorithm[J]. European Journal of Operational Research, 2018, 271(3): 808-817.

[17] Kingma D P. Adam: A method for stochastic optimization[J]. arxiv preprint arxiv: 1412. 6980, 2014.

[18] Fang W, Yao X, Zhao X, et al. A stochastic control-approach to maximize profit on service provisioning for mobile cloudlet platforms[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2016, 48(4): 522-534.

[19] Ward R, Wu X, Bottou L. Adagrad stepsizes: Sharp convergence over nonconvex landscapes[J]. The Journal of Machine Learning Research, 2020, 21(1): 9047-9076.

[20] Kurbiel T, Khaleghian S. Training of deep neural networks based on distance measures using RMSProp[J]. arXiv preprint arXiv: 1708. 01911, 2017.

[21] Ding J, Ren X, Luo R, et al. An adaptive and momental bound method for stochastic learning[J]. arXiv preprint arXiv: 1910. 12249, 2019.

[22] Abadi M, Chu A, Goodfellow I, et al. Deep learningwith differential privacy[C]//Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2016: 308-318.