吉彦卿,张玉金,徐昊君. 基于联邦平均的差分隐私算法设计与实现[J]. 智能计算机与应用,2024,14(6):201-206. DOI:10.20169/j.issn.2095-2163.240630

基于联邦平均的差分隐私算法设计与实现

吉彦卿, 张玉金, 徐昊君
(上海工程技术大学 电子电气工程学院, 上海 201620)

摘 要: 随着信息技术的快速发展,各行业领域的数据大量涌现,以数据驱动的机器学习模型已广泛应用于人们的日常工作和生活中。准确性和泛化能力是机器学习模型的 2 个重要指标,这需要从大量有用的数据中进行学习。在一些机器学习应用场景中,各个数据来源方无法直接进行数据交换,严重影响了机器学习模型的能力提升。针对数据隐私泄露风险问题,本文设计并实现了一种基于联邦平均的差分隐私保护算法,兼顾学习训练效率和数据隐私保护能力。实验结果表明,本文算法在牺牲了一部分通信效率的情况下,可达到较好的数据隐私保护能力。

关键词: 联邦学习; 机器学习; 数据隐私; 数据驱动
中图分类号: TP309 **文献标志码**: A **文章编号**: 2095-2163(2024)06-0201-06

Design and implementation of differential privacy algorithm based on federated average

JI Yanqing, ZHANG Yujin, XU Haojun
(School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

Abstract: With the rapid development of information technology, a large amount of data has been generated, and data-driven machine learning models have been widely applied in people´s daily work and life. Accuracy and generalization ability are two important indicators of machine learning models, which require learning from a large amount of useful data. In some machine learning application scenarios, various data sources are unable to directly exchange data, which seriously affects the ability improvement of machine learning models. In response to the risk of data privacy leakage, this paper designs and implements a differential privacy protection algorithm based on federated average, which combines learning and training efficiency with data privacy protection capabilities. The experimental results show that the proposed algorithm can achieve good data privacy protection capabilities while sacrificing some communication efficiency.

Key words: federated learning; machine learning; data privacy; data-driven

0 引 言

机器学习是人工智能的核心研究领域。传统大规模的机器学习模型[1]是将所有待处理、待训练的数据收集到本地进行操作,然而科技的发展也带来更多的挑战,例如行业内的竞争关系和行业间的数据隐私保护需求,致使将各方的数据整合起来比较困难。由于数据被赋予了“资产”的价值,个人、公司和政府三方都越来越重视数据的隐私权和所有权。因此,人工智能发展面临着“数据孤岛”的问题仍亟待解决。为了应对这一问题,联邦学习则应运而生,这是一种拥有隐私保护能力的分布式机器学习算法,能够有效降低用户隐私泄露风险,提高系统的抗攻击性能。

与现有的分布式学习相比,联邦学习的优势主要体现在:联邦学习把数据分布在各参与方,实现“模型动数据不动”和“数据可用不可见”,一定程度上保护了参与方原始数据的安全性。然而,现有方法无法保证每个联邦学习参与方都是完全可信赖的,有些意外的情况仍然可能发生,例如有一个或多个参与方是恶意的,出于某些目的,会使用推测法将联邦学习模型训练的参数反向推理出其他联邦学习参与方的原始输入数据,造成了严重的隐私泄露

作者简介: 吉彦卿(1996-),男,硕士研究生,主要研究方向:多媒体内容安全; 徐昊君(2000-),男,本科生,主要研究方向:机器学习。
通讯作者: 张玉金(1982-),男,博士,副教授,主要研究方向:多媒体内容安全,图像处理,模式识别。Email:yjzhang@ sues.edu.cn

收稿日期: 2023-04-14

后果。

为了避免或对抗恶意参与方反向推测其他参与方原始数据,近年来一些研究学者将联邦学习与差分隐私进行结合[2],差分隐私不仅对用户提供了更强的隐私保护支持,还为模型提供了有力的数学理论基础。Wang 等学者[3]提出了一种基于本地差分隐私(Local Differential Privacy)的多维数据训练模型,建立了一种新的本地差分隐私和随机梯度下降的协同模型。Li 等学者[4]分析了联邦平均算法(Federated Averaging Algorithm, FedAvg)在非独立同分布数据上的收敛性,建立了收敛速度对于强凸和光滑问题,并证明了通信效率和收敛速度之间的平衡。由于用户设备可能断连,将设备全部参与的假设放宽为部分设备参与,可以在不严重减慢学习速度的情况下实现低设备参与率。Wu 等学者[5]提出了一种优化联邦学习而不会破坏数据隐私的新算法,自适应地对每个用户进行评估并调整其参与概率,以减轻低价值客户对培训过程的影响,算法的选择性行为导致通信轮数和全局模型收敛的时间量显著降低。

在综合分析前人工作的基础上,为了兼顾学习训练效率和数据隐私保护能力,本文设计并实现了一种基于联邦平均的差分隐私保护算法。

1 基于联邦平均的差分隐私算法

1.1 算法设计

联邦平均框架示意如图 1 所示。由图 1 可知,联邦平均框架的基本步骤如下:

步骤 1 假设一共有 k 个客户端(worker),客户端 1、客户端 2、……、客户端 m 分别表示每个被选择的客户端的数据集。如标号①所示,客户端本地训练机器学习模型;

步骤 2 客户端 m 与中央服务器交互通信的第 t 次,中央服务器从 m 个被选定的客户端中随机任意选择组成集合,被选定的集合中的客户端做本地梯度降低任务,然后模型更新参数上传给服务器,注意此过程只上传模型参数而不上传原始数据,如标号②所示;

步骤 3 中央服务器聚合从客户端收集到的模型更新参数,并更新其模型参数,服务器聚合模型更新参数的目的在于综合考虑每个客户端的模型参数,以便对全局共享模型进行优化,如标号③所示;

步骤 4 中央服务器将聚合后的模型参数 w_t 发送给客户端,并开始下一轮训练过程更新模型。如标号④所示。

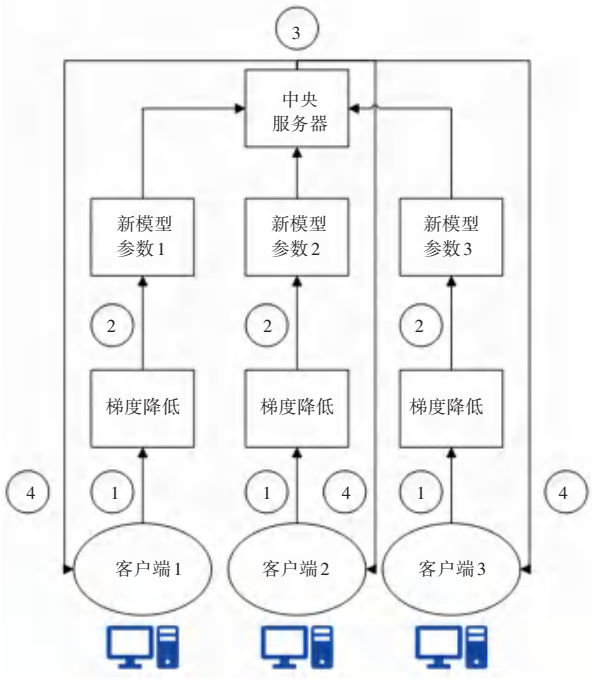

图 1 联邦平均框架示意图

Fig. 1 Illustration of federal average framework

这里,将对联邦平均算法步骤[6]展开阐释分述如下。

算法 1 联邦平均算法

k 表示 K 个用户端的第 k 个用户端($k \in K$);B 表示本地批量数据集的大小,E 表示本地轮数的数量,η 表示选择的学习率。

服务端执行事务:

1.初始化参数 w_0// 设 w_0 为初始化模型参数

2.for 每轮 $t = 1,2,\cdots$ 执行

将 $\max(C \cdot K,1)$ 赋值给 m

从 m 个被选定的用户端中随机组成 S_t 集合

for 每个 S_t 集合中的用户端并行计算

将第 k 个用户的用户更新 (k,w_t) 赋值给 w_{t+1}^k

将 $\sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$ 赋值给 w_{t+1} //聚合用户端更新模型参数为 w_{t+1}

用户端执行事务:

将本地的 k 个用户端的数据库大小赋值给 B

for 每轮 i 从 1 到 E 执行

for 本地批量数据 $b \in B$ 执行

将 $w - \eta \tilde{N}\ell(w;b)$ 赋值给 w

将 w 回传给服务端

考虑到差分隐私随机梯度下降算法受通信效率限制，本文使用联邦平均算法来对其进行优化。与联邦平均算法一样，联邦平均差分隐私分为用户端和服务端两类。用户端和服务端的工作有一定区别，其中用户端主要为本地计算梯度降低和进行参数裁剪，服务端主要是聚合参数模型、裁剪梯度、更新全局模型参数。以设置添加高斯噪声为例，基于差分隐私的随机梯度下降算法[7]的描述如下。

算法 2　基于差分隐私的随机梯度下降算法

输入　训练样本数据 $\{x_1,\cdots,x_N\}$；损失函数：$\frac{1}{N}\sum_i \mathcal{L}(\theta,x_i)$。模型参数：学习率 η_t，梯度裁剪边界 C，噪声标准差 σ，每一轮训练样本大小 L

输出　模型参数 θ_T 以及计算全局差分隐私成本

1. 随机地初始化模型参数 θ_0

2. for 对每轮的 $t \in [1,T]$ 执行以下梯度更新

从样本集中随机挑选大小为 L 的集合，选出 L_t 的概率是 L/N

for 对每个样本 $i \in L_t$，计算 $\boldsymbol{g}_t(x_i) \leftarrow \tilde{\nabla}_{\theta_t}\mathcal{L}(\theta_t,x_i)$；

裁剪梯度 $\tilde{\boldsymbol{g}}_t \leftarrow \frac{1}{L}\left(\sum_i \bar{\boldsymbol{g}}_t(x_i) + N(0,\sigma^2C^2\boldsymbol{I})\right)$

噪声 $\tilde{\boldsymbol{g}}_t \leftarrow \frac{1}{L}\left(\sum_i \bar{\boldsymbol{g}}_t(x_i) + N(0,\sigma^2C^2\boldsymbol{I})\right)$

得到梯度降低后的模型参数 $\theta_{t+1} \leftarrow \theta_t - \eta_t\tilde{\boldsymbol{g}}_t$

本文算法用户端工作流程的描述如下。

算法 3　联邦平均差分隐私算法用户端

输入　参数包括用户端的 id：k；全局模型 θ^0；模型参数：学习率 η；本地迭代 E 次；训练样本大小 B

输出　模型更新：$\Delta_k = \theta - \theta^0$

1.服务端发出全局训练模型参数 θ^0，然后更新本地模型 $\theta = \theta - \theta^0$

2.for 每一轮 $i = 1,2\cdots\cdots,E$，执行以下操作

将本地数据分割为一共 $|B|$ 份的 B 数据

for 对每个批次 $b \in B$

梯度降低：$\theta \leftarrow \theta - \eta\tilde{\nabla}\ell(\theta;b)$；

裁剪模型参数：$\theta \leftarrow \theta^0 + ClipFn(\theta - \theta^0)$

由此可知，从所有的用户端中选定一个集合 C_t 来参与学习训练，然后服务端对被选定的用户端发出命令，执行本地模型训练。服务端的工作流程描述如下。

算法 4　联邦平均差分隐私算法服务端

输入　训练的样本数据：$\{x_1,\cdots,x_N\}$；损失函数：$l(\theta)=\frac{1}{N}\sum_{i=1}^{N} l(\theta;x_i)$；学习率 η_t，梯度裁剪边界 C；噪声参数：σ；训练样本大小 B

输出　全局模型参数 θ^{t+1}

1.随机地初始化训练模型参数 θ^0

2.定义客户端的权重：第 i 个用户端 c_i 的权重是 $w_k = \min\left(\frac{n_k}{\hat{w}},1\right)$

3.设 $W = \sum_1^K w_k$

4.for 对每一轮的 $t = 1,2\cdots\cdots$

5.　任意挑选参与训练的本轮训练的用户端集合 C_t

6.　for 对每个用户 $k \in C_t$

本地梯度降低：$\Delta_k^{t+1} \leftarrow UserUpdate(k,\theta^t,\text{ClipFn})$

加权平均聚合用户端参数：

$$\Delta^{t+1} = \begin{cases} \dfrac{\sum_{k\in C^t} w_k\Delta_k}{qW}, & \text{for } \tilde{f}_f \\ \dfrac{\sum_{k\in C^t} w_k\Delta_k}{\max(qW_{\min},\Sigma_{k\in C^t}w_k)}, & \text{for } \tilde{f}_c \end{cases}$$

裁剪 Δ^t：$\Delta^t \leftarrow \Delta^t/\max\left(1,\frac{\|\Delta^t\|}{C}\right)$

求高斯噪声方差

$$\sigma \leftarrow \left\{\frac{zS}{qW}\text{ for }\tilde{f}_f \text{ or } \frac{2zS}{qW_{\min}}\text{ for }\tilde{f}_c\right\}$$

更新全局模型参数 $\theta^{t+1} \leftarrow \theta^t + \Delta^{t+1} + N(0,I\sigma^2)$

由此可知，服务端接受所有被挑选的用户端计算好的模型参数 Δ_k^t，并进行加权平均聚合，得到 Δ^t。之后求高斯噪声分布的方差，根据高斯分布 $N(0,I\sigma^2)$ 生成噪声数据。在全局模型聚合的过程中，添加生成的噪声数据，相加得到新的全局模型参数 θ^t。最后不断重复迭代上述步骤，直到模型收敛于某一个最优值为止。

1.2　算法分析

基于联邦平均的差分隐私和差分隐私随机梯度下降算法相比，具有通信效率方面的优势，这是因为联邦平均的特点是客户端和服务端分工协作，根据

算法中 $\Delta_k^{t+1} \leftarrow UserUpdate(k,\theta^t,ClipFn)$ 这一步,证明客户端负责在本地计算梯度降低,并将计算好的梯度参数传给服务端,而服务端只负责将传过来的梯度参数进行加权平均模型聚合,然后更新全局模型参数,再分发全局模型参数。但是差分隐私随机梯度下降算法仅有差分隐私的安全性,然而当用户端数量增长到一定量级时,且当数据集为非独立同分布时,以及当网络通信差、有些设备掉线时,联邦平均方法明显优于差分隐私随机梯度下降的效果,这是因为差分隐私随机梯度下降是假设数据集独立同分布、用户端数量也不庞大、网络通信较好的情况。但实际的情况较复杂,联邦平均比差分隐私方法具有更好的适应性。

2 实验结果与分析

为了验证本文方法的有效性,使用 Pycharm 集成开发环境和 Pytorch 联邦学习框架进行模型搭建,实验数据集选取 Mnist 数据集,包含 250 个不同人的手写数字图像,分为测试集和训练集,其中测试集包括 10 000 幅手写数字图像,训练集包括 60 000 幅手写数字图像。联邦学习参数和设置如下:用户端 K 为 100 个,随机挑选占比 C 为 0.1,本地迭代次数 E 为 5 次,学习率 0.01,剩余参数默认值。深度网络模型设置:选取卷积神经网络作为分类网络,包括 2 个卷积层和 2 个全连接层,卷积核大小选取为 5。在卷积运算后,接着执行 Max pooling 池化操作,选取 *ReLU* 函数作为神经元的激活函数。为了提升网络的学习效率,采用 *Dropout* 方法抛弃弱连接权值。下面从通信效率和隐私保护参数两个方面对模型性能进行具体分析。

2.1 通信效率

本节分析联邦平均和联邦平均差分隐私实验结果,无噪声方式 50 轮的测试准确率如图 2 所示。对照组数据同分布训练 50 轮和 300 轮,除了差分隐私方式采用添加高斯噪声方式,其他所有参数均一样,结果如图 3 和图 4 所示。通过数据对比可知,图 2 的准确率可达 85.80%,已经取得了较高的准确率,而图 3 的准确率只有 56.00%,训练了 50 轮只能达到无添加噪声方式的第 19 轮训练的准确率。分析可知,低的准确率完全无法满足实际需要。如果要达到无噪声训练 50 轮的准确率,需要继续进行训练,根据对照组 300 轮训练的数据情况可推知,在训练第 300 轮后,添加高斯噪声方式的准确度达到了 85.26%,已经和无噪声方式训练 50 轮的准确度相差不到 1%。因此,在联邦平均训练中添加噪声的行为会造成模型训练准确度降低,意味着需要进行更多轮的训练来弥补下降的准确度,所以耗费在训练上的时间成本增加,相应的通信成本也有所提高。

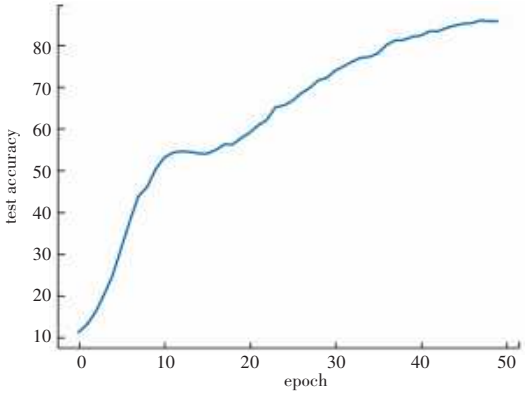

图 2 无噪声方式 50 轮的测试准确率

Fig. 2 Accuracy rate of 50 rounds of noise-free test method

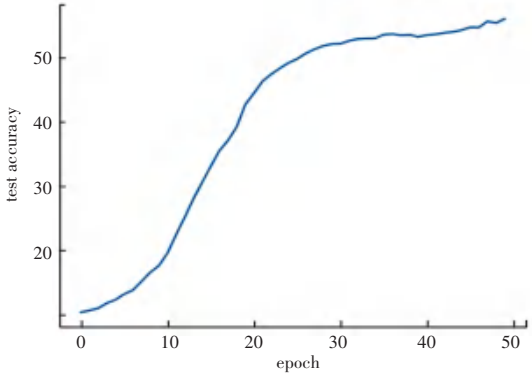

图 3 添加高斯噪声方式 50 轮的测试准确率

Fig. 3 Test accuracy of 50 rounds with added Gaussian noise

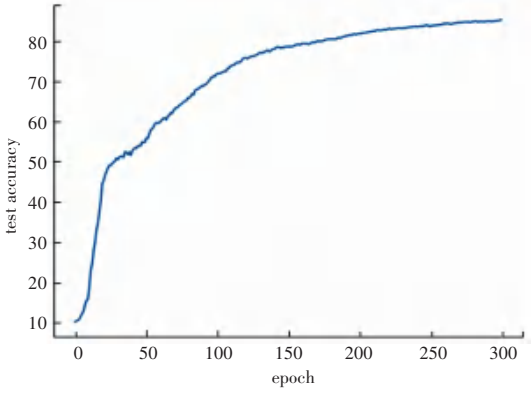

图 4 添加高斯噪声方式 300 轮的测试准确率

Fig. 4 Test accuracy of 300 rounds with added Gaussian noise

2.2 隐私保护参数

参数 δ 和参数 ε 的变化也会影响最终的准确率。为了评估参数 δ 和参数 ε 对模型产生的影响,实验统一使用添加高斯噪声的方式,首先,固定参数

δ 不变，改变参数 ε，观察准确率如何变化。实验的参数 δ 都固定为 0.01，训练轮数都固定为 500 轮。研究中当将参数 ε 分别设置为 6、8、10 时，实验结果分别如图 5、图 6 和图 7 所示。

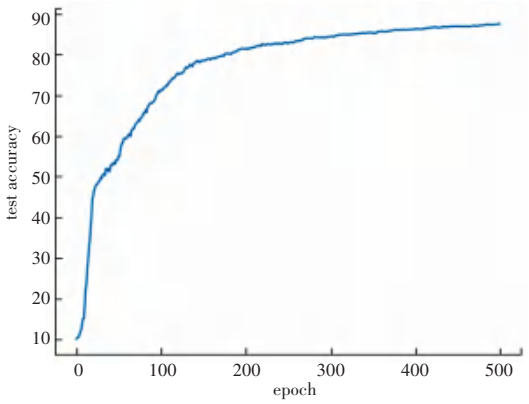

图 5　准确率变化曲线（ε=6）

Fig. 5　Accuracy curve（ε=6）

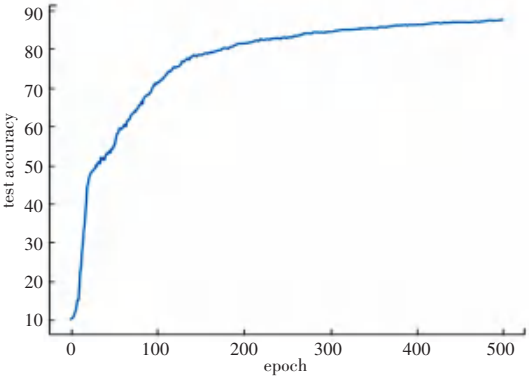

图 6　准确率变化曲线（ε=8）

Fig. 6　Accuracy curve（ε=8）

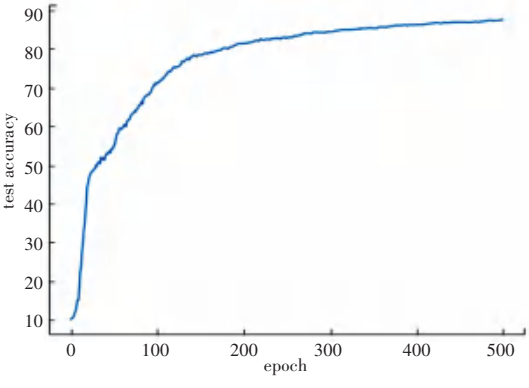

图 7　准确率变化曲线（ε=10）

Fig. 7　Accuracy curve（ε=10）

由图 5～图 7 分析可知，当参数 δ 固定不变，参数 ε 分别设置为 6、8、10 时，准确率分别为 87.05%、87.55%、87.85%。尽管三者准确率相差不足 1%，但仍可以发现：即当参数 ε 设定越大时，准确率越高。另一方面，由于参数 ε 越接近 0 则表示噪声的差分隐私保护性越强，准确率提高、隐私保护能力却会降低。同时，三者准确率初次超过 80% 的训练轮数，分别是第 186、176、176 轮，可知参数 ε 越大，准确率收敛的速度越快。此外，无论参数 ε 设置为 8 或 10，准确率初次超过 80% 的训练轮数都是第 176 轮，说明收敛速度已经接近最优。

当参数 ε 固定为 10，改变参数 δ，训练轮数固定为 500 轮，参数 δ 分别设置为 0.1、0.001、0.000 01，实验结果分别如图 8～图 10 所示，进一步可以得到准确率分别为 88.21%、87.59% 和 87.24%。根据图 8～图 10 数据可知，将参数 ε 固定，当参数 δ 设定越大时，准确率越高。同时，分别列出三者准确率初次超过 80% 的训练轮数，分别是第 169、176、185 轮，因此可推出参数 δ 越大，准确率收敛的速度越快。但当参数 ε 固定时，参数 δ 越接近 0，模型的隐私保护能力越强，因此这反映了一个规律，差分隐私在增强模型隐私保护能力的同时，必定会降低收敛速度，间接降低了通信效率。

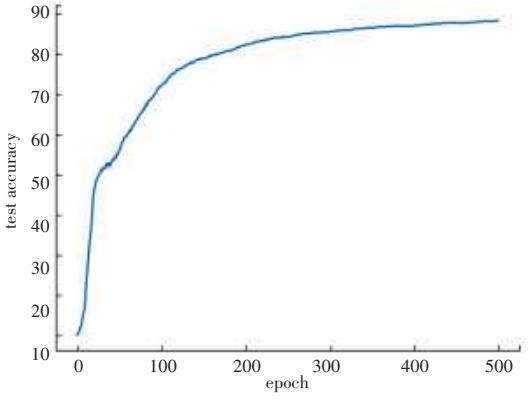

图 8　准确率变化曲线（δ=0.1）

Fig. 8　Accuracy curve（δ=0.1）

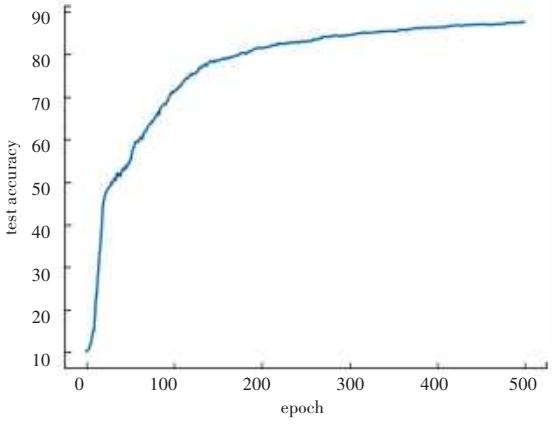

图 9　准确率变化曲线（δ=0.001）

Fig. 9　Accuracy curve（δ=0.001）

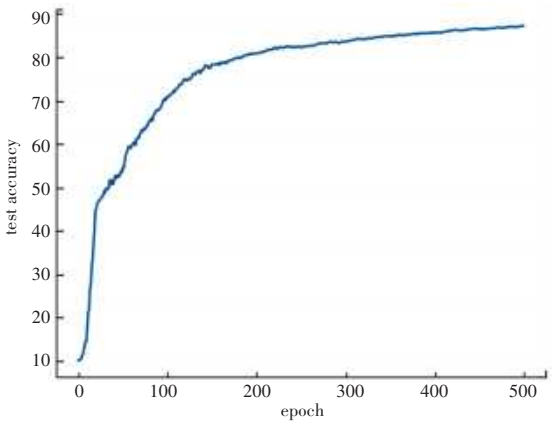

图 10 准确率变化曲线(δ=0.000 01)

Fig. 10 Accuracy curve(δ=0.000 01)

从上述实验结果可知,联邦学习差分隐私拥有隐私保护能力,通过合理调整参数 ε、参数 δ 可以在保证隐私保护能力的同时,尽量不过多降低通信效率。现实中如果需求是对于期望隐私保护能力强,可以放弃一部分通信效率,可以把参数 ε 和参数 δ 适当调低。如果需求是对于通信效率要求高,可以放弃一部分隐私保护能力,应当把参数 ε 和参数 δ 适当调高。

3 结束语

针对数据隐私泄露风险问题,本文设计并实现了一种基于联邦平均的差分隐私保护算法,该算法兼顾学习训练效率和数据隐私保护能力,实验结果验证了该算法较好的数据隐私保护能力。未来可以深入研究如何优化通信效率和隐私保护能力。

参考文献

[1] 王志文,刘广起,韩晓晖,等. 基于机器学习的恶意软件识别研究综述[J]. 小型微型计算机系统,2022,43(12):2628-2637.

[2] 于梦晴. 面向联邦学习的两种类型差分隐私研究[D]. 杭州:浙江科技学院, 2021.

[3] WANG Zhibo, SONG Mengkai, ZHANG Zhifeng, et al. Beyond inferring class representatives: User-level privacy leakage from federated learning[C]//IEEE INFOCOM 2019-IEEE Conference on Computer Communications. Paris, France: IEEE, 2019: 2512-2520.

[4] LI Xiang, HUANG Kaixuan, YANG Wenhao, et al. On the convergence of fedavg on non-IID data[J]. arXiv preprint arXiv: 1907.02189v4,2020.

[5] WU Wentai, HE Ligang, LIN Weiwei, et al. FedProf: Selective federated learning with representation profiling[J]. arXiv preprint arXiv:2102.01933v9,2022.

[6] MCMAHAN H B, MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data[J]. arXiv preprint arXiv:1602.05629 v3, 2017.

[7] 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. Vienna, Austria:ACM, 2016: 308-318.