面向非独立同分布数据的自适应联邦深度学习算法

张泽辉 [1] 李庆丹 [1] 富 瑶 [1] 何宁昕 [1] 高铁杠 [1]

摘 要 近些年, 联邦学习 (Federated learning, FL) 由于能够打破数据壁垒, 实现孤岛数据价值变现, 受到了工业界和学术界的广泛关注. 然而, 在实际工程应用中, 联邦学习存在着数据隐私泄露和模型性能损失的问题. 为此, 首先对这两个问题进行数学描述与分析. 然后, 提出一种自适应模型聚合方案, 该方案能够设定各参与者的 Mini-batch 值和自适应调整全局模型聚合间隔, 旨在保证模型精度的同时, 提高联邦学习训练效率. 并且, 混沌系统被首次引入联邦学习领域中, 用于构建一种基于混沌系统和同态加密的混合隐私保护方案, 从而进一步提升系统的隐私保护水平. 理论分析与实验结果表明, 提出的联邦学习算法能够保证参与者的数据隐私安全. 并且, 在非独立同分布数据的场景下, 该算法能够在保证模型精度的前提下提高训练效率, 降低系统通信成本, 具备实际工业场景应用的可行性.

关键词 联邦学习, 深度学习, 隐私保护, 同态加密, 混沌系统

引用格式 张泽辉, 李庆丹, 富瑶, 何宁昕, 高铁杠. 面向非独立同分布数据的自适应联邦深度学习算法. 自动化学报, 2023, **49**(12): 2493−2506

DOI 10.16383/j.aas.c201018

Adaptive Federated Deep Learning With Non-IID Data

ZHANG Ze-Hui[1] LI Qing-Dan[1] FU Yao[1] HE Ning-Xin[1] GAO Tie-Gang[1]

Abstract In recent years, federated learning (FL) that can break data barriers and realize the value of isolated data, has been received wide-spread attention from industry and academia. However, in real industry applications, federated learning has problems with privacy leakage and model accuracy loss, which is analyzed through mathematical demonstration in this study. To solve the issues, this paper proposes an adaptive global model aggregation scheme that can adaptively set the Mini-batch value of each participant and the global model aggregation interval for the parameter server, which aims to improve the training efficiency while ensuring the accuracy of the model. Moreover, this paper introduces the chaos system into the federated learning field, which is used to construct a hybrid privacy-preserving scheme based on chaos system and homomorphic encryption, thereby further improving the privacy protection level. Theoretical analysis and experimental results show that the proposed approach can guarantee the data privacy security of participants. Moreover, in the non-independent and identically distributed (Non-IID) data scenario, the proposed method can improve the training efficiency and reduce communication cost while ensuring the model accuracy, which is feasible for real industrial applications.

Key words Federated learning (FL), deep learning, privacy-preserving, homomorphic encryption, chaos system

Citation Zhang Ze-Hui, Li Qing-Dan, Fu Yao, He Ning-Xin, Gao Tie-Gang. Adaptive federated deep learning with non-IID data. *Acta Automatica Sinica*, 2023, **49**(12): 2493−2506

近些年, 以深度学习算法为代表的人工智能技术在故障诊断、智能控制、情感识别以及生物信息等领域得到了广泛应用[1−4], 而决定深度学习模型性能的关键之一是高质量的训练数据. 目前构建深度学习模型通常采用的是传统集中式学习方法 (Centralized learning, CL). 在该方法中, 训练者提前收集各方的数据集, 然后进行数据聚合、预处理等操作, 最后使用处理后的数据进行深度学习模型的训练. 然而, 在某些行业 (例如医疗行业), 由于数据隐私等相关的法律规定, 数据拥有者通常不愿意分享数据[5−8].

为打破数据壁垒, 联邦学习 (Federated learning, FL) 于 2015 年由 Google 公司提出[9]. 该方法能够组织不同参与者通过分享其本地模型参数, 协作训练一个全局深度学习模型[10]. 由于联邦学习系统中的参与者不需要分享本地数据, 因此该技术十分适用于某些对数据敏感的工业场景. Zhang 等[11] 针对真实场景中存在的设备故障数据难收集的问

收稿日期 2020-12-08 录用日期 2021-03-19
Manuscript received December 8, 2020; accepted March 19, 2021
国家科技重大专项基金 (2018YFB0204304), 天津市研究生科研创新基金 (2019YJSB067) 资助
Supported by National Science and Technology Major Project of China (2018YFB0204304) and Tianjin Research Innovation Project for Postgraduate Students (2019YJSB067)
本文责任编委 杨健
Recommended by Associate Editor YANG Jian
1. 南开大学软件学院 天津 300071
1. College of Software, Nankai University, Tianjin 300071

题, 提出了一种基于联邦学习的轴承故障诊断方法. 该方法通过多个参与者对故障诊断模型进行本地训练, 从而构建一个高性能的全局故障诊断模型. Sheller 等[12] 指出联邦学习在医疗场景的巨大潜力, 并针对病人健康数据的隐私保护需求, 提出了一种基于联邦学习的医学诊断模型. 该模型能够跨多个研究机构进行医学诊断模型的训练, 从而提高诊断模型的性能. Kwon 等[13] 针对海洋场景中无线传感器通信困难等问题, 首次将联邦学习算法应用到智能海洋网络的构建. 通过以上应用研究可以看出, 联邦学习算法在工程领域应用有着广阔前景.

为了进一步提高联邦学习的性能, Rothchild 等[14] 提出使用梯度选择和自适应调整学习速率方法来提升联邦学习训练效率, 降低通信代价. Duan 等[15] 利用数据增强技术减轻由于数据分布不均造成的联邦学习模型性能的衰减, 从而提高联邦学习模型的性能. Liu 等[16] 将动量项引入到联邦学习中, 提出一种动量联邦学习算法. 在该算法中, 参与者的本地训练中使用动量梯度下降算法来加速模型收敛. Wang 等[17] 提出了一种自适应联邦学习算法, 能够在固定的资源预算下自适应调整模型聚合频率. 该算法是以算法经验损失函数为凸模型为前提设计的, 对于传统机器学习算法 (例如支持向量机、线性回归和 K-Means 算法) 在联邦学习框架下的应用, 能够比较好地实时估算全局模型聚合频率. 然而, 对于经验损失函数为非凸模型, 例如深度神经网络、卷积神经网络和递归神经网络等, 该算法难以计算出最佳模型聚合频率.

以上研究都未对参与者的数据隐私安全方面进行考虑. 近期的研究[18−19] 指出, 参与者上传的模型参数有可能泄露本地的数据隐私信息. 为保护参与者的隐私信息, 多种面向机器学习的数据隐私保护方法相继提出, 主要可以分为两类: 基于差分隐私的方法和基于同态加密的方法.

1) 基于差分隐私的方法: 差分隐私 (Differential privacy, DP) 是一种常见的数据隐私保护技术[20]. 它采用某种随机机制 (例如随机采样或添加噪声) 来对用户处理的输入或输出数据进行扰动, 从而使用户处理的结果在一定程度上可以对抗隐私分析. Gong 等[7] 提出了一种基于相关性的自适应差分隐私算法, 该算法根据不同层神经元之间的相关性对模型的梯度参数施加扰动, 从而保护用户数据的隐私信息. Wang 等[21] 提出通过对模型的输入特征数据注入噪声, 从而为众包数据提供差分隐私保护, 并估计与目标类别相关的特征重要性, 遵循较少的噪声注入原则以确保模型的有效性. 基于差分隐私的隐私保护算法具有计算复杂度低和易于实现使用等优点, 但是会导致模型的性能下降.

2) 基于同态加密的方法: 同态加密支持对密文进行数学运算, 密文计算的结果经解密后与明文计算结果相同[6, 22], 该特性十分适用于构建支持隐私保护的机器学习算法. 宋蕾等[23] 提出通过对训练数据集同态加密, 进而构建支持隐私保护的联合逻辑回归学习算法. Phong 等[24] 提出对参与者上传的梯度参数进行同态加密, 从而保护联邦学习中参与者的隐私信息. 张泽辉等[19] 提出对参与者上传模型权重参数进行同态加密, 从而保护各参与者的隐私信息. 同态加密算法有着很好的安全性, 并且支持构建高精度的深度学习模型, 但在训练过程中增加的额外加密/解密运算会消耗大量的计算资源和训练时间. 因此, 某些研究[19] 采用较低的全局模型更新频率, 即较大的模型聚合间隔 τ, 以减少加密/解密次数和通信成本, 从而提高联邦学习训练效率. 然而, 在非独立同分布数据场景下, 较低的更新频率会导致联邦学习模型的性能大幅降低, 甚至直接导致模型不收敛. 反之, 采用较高全局模型更新频率[24], 又会极大地增加联邦学习训练的通信成本和计算代价.

在工业场景中, 模型的精度对生产效益等方面有着巨大影响. 因此, 本文选用同态加密方案构建支持隐私保护的联邦学习算法. 针对联邦学习系统中存在的性能优化与隐私保护问题, 本文提出了一种面向非独立同分布数据的自适应联邦深度学习算法. 本文研究工作主要如下:

1) 针对联邦学习算法在工业场景中的应用, 通过数学推导分析总结了联邦学习模型出现精度损失的主要原因. 基于此, 本文首次提出面向联邦深度学习的自适应全局模型聚合方案, 该方案能够设定各参与者的 Mini-batch 值, 并根据参与者训练过程中的信息, 自适应调整全局模型聚合频率, 在保证模型性能的前提下, 提高联邦学习训练效率.

2) 借鉴经典图像加密方法, 本文首次将混沌系统引入联邦学习领域中, 并采用 CKKS (Cheon-Kim-Kim-Song) 同态加密方案构建出一种基于混沌系统和同态加密的混合隐私保护方案. 与基于 Paillier 同态加密的联邦学习算法相比, 本文所提出的联邦学习算法大大提高了模型的训练速度.

3) 理论分析和实验结果表明, 本文提出的方法能够保证参与者的数据隐私安全. 并且, 在非独立同分布数据的场景下, 该方法能够在保证模型精度的前提下提高训练效率, 降低系统通信成本, 具备实际工业场景应用的可行性.

1　预备知识

1.1　同态加密

同态加密算法具有的同态特性是指, 在密文上执行某种基础数学运算, 将得到的密文结果解密后与原明文进行相同的数学操作, 结果相同. 本文选用的是 CKKS 同态加密算法, 该算法相比于传统的 Paillier 加密算法有着更快的加密/解密速度[25]. 该方案简述如下 (详细方案可见文献 [26–27]).

1) 密钥生成 $keygen(1^{\lambda})$: 输入安全参数 λ, 选择整数 p, L, 设定 $q_{\ell}=p^{\ell}$, $\ell=1,\ 2,\cdots,L$. 算法的最终输出为 $(sk,\ pk,\ evk)$.

2) 编码算法 $Encode(\boldsymbol{z};\ \Delta)$: 对于一个 $(N/2)$ 维的高斯整数向量 $\boldsymbol{z}=(z_j)_{j\in T}\in\mathbf{Z}[i]^{N/2}$, 计算 $\lfloor\Delta\pi^{-1}(\boldsymbol{z})\rfloor_{\sigma(\mathbf{R})}$. 输出其在标准嵌入图中的逆.

3) 解码算法 $Decode(m;\ \Delta)$: 对于输入多项式 $m(X)\in\mathbf{R}$, 计算出对应的向量 $\pi\circ\delta(m)$ 最终输出:

$$\boldsymbol{z}=(z_j)_{j\in T}\in\mathbf{Z}[i]^{\frac{N}{2}} \tag{1}$$

4) 加密 $Encrypt\ (m\in\mathbf{R})$: 取 $v\leftarrow\mathcal{ZO}(0.5)$ 和 $e_1,\ e_2\leftarrow\mathcal{DG}(\sigma^2)$, 对明文 m 加密运算为

$$[v\cdot pk+(m+e_1,\ e_2)]q_{\mathcal{L}}\in\mathbf{R}_{q_{\mathcal{L}}}^2 \tag{2}$$

5) 解密 $Decrypt\ (ct\in\mathbf{R}_{q_{\ell}}^2;sk)$: 对于密文 ct, 解密运算为

$$m=[\langle ct,\ sk\rangle]\,q_{\ell} \tag{3}$$

6) 密文加法 $Add\ (ct_1\in\mathbf{R}_{q_{\ell}}^2,\ ct_2\in\mathbf{R}_{q_{\ell}}^2)$: 对于两个密文 ct_1 和 ct_2 输出为

$$ct_{\text{add}}=[\langle ct_1,\ ct_2\rangle]q_{\ell} \tag{4}$$

式 (1) ~ (4) 中, $\mathcal{DG}(\sigma^2)$ 代表从 $\mathbf{Z}^N$ 中生成一个 N 维多项式向量, 该向量中每个系数服从方差为 σ^2 的离散高斯分布, $\mathcal{ZO}(\rho)$ 代表从 $\{-1,\ 0,\ 1\}^N$ 中生成一个 N 维向量, 其中生成 0 的概率为 $(1-\rho)$, 而生成 -1 和 1 的概率为 $\rho/2$.

1.2　混沌系统

混沌系统具有敏感性、非周期性和不可预测性等特点, 常用于生成伪随机序列[28]. Robert Matthews 提出的 Logistic 映射, 已经广泛用于文本加密、图像加密以及视频加密等领域[29–30]. 因此, 本文将使用混沌 Logistic 映射系统生成伪随机数列, 其数学表达式为

$$x(i+1)=\lambda\cdot x(i)\cdot(1-x(i)) \tag{5}$$

式中, $x(i)\in(0,\ 1)$, 并且 $0<\lambda\leq 4$. 当 $3.56994<\lambda\leq 4$ 时, Logistic 映射进入混沌状态.

文献 [31] 提出参与者采用相同的置乱方式对模型参数进行加密, 从而起到保护参与者数据隐私的作用. 因此, 本文采用该图像置乱加密方法[28] 对参与者模型参数进行加密, 当采用推理的方法 (具体见第 2.1 节) 对置乱的模型参数进行分析时, 所获得的数据是经过置乱加密的数据, 从而能够一定程度地保护好参与者的数据隐私. 置乱加密过程为: 参与者使用混沌系统生成与局部模型权重参数数量相同的随机数列 R_{chaos}, 然后对 R_{chaos} 进行升序排列, 得到其索引序列 $Index$, 再按照索引序列 $Index$ 的值, 对参与者的模型权重参数 $w_{\text{par},i}$ 进行置乱. 特别地, 各参与者使用混沌系统的初始状态参数相同, 从而保证参数服务器的模型聚合正确性.

2　联邦学习信息泄露与性能损失

2.1　隐私泄露

联邦学习在训练过程中, 参与者分享的本地模型参数可能会泄露其本地的数据隐私信息. 图 1 为一个典型的 3 层神经网络 (隐含层使用 Dropout 技术). $\boldsymbol{x}^{(0)}$ 为神经网络的输入数据, y 为神经网络的目标输出数据.

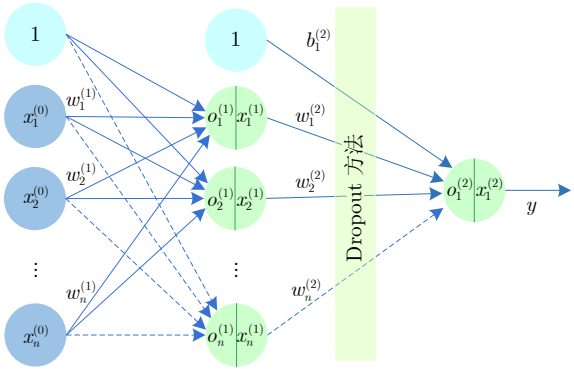


图 1　多层神经网络模型

Fig. 1　Multi-layer neural network model

1) 情况 1 (不考虑 Dropout)

首先求出模型输出值与目标值的 loss 值, 计算式为

$$L=\frac{1}{2}\left(x_1^{(2)}-y\right)^2 \tag{6}$$

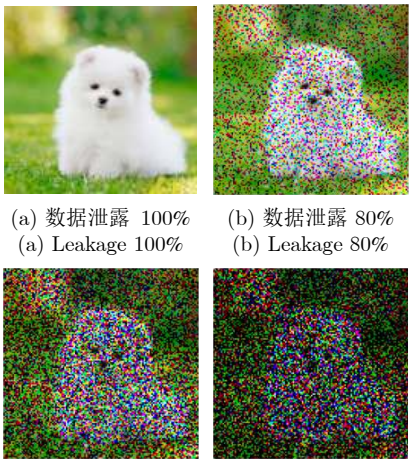
然后使用链式求导法则, 可以求得

$$\frac{\partial L}{\partial w_1^{(1)}}=\frac{\partial L}{\partial x_1^{(1)}}\frac{\partial x_1^{(1)}}{\partial o_1^{(1)}}\frac{\partial o_1^{(1)}}{\partial w_1^{(1)}}=\frac{\partial L}{\partial x_1^{(1)}}x_1^{(1)}\left(1-x_1^{(1)}\right)x_1^{(0)} \tag{7}$$

$$\frac{\partial L}{\partial b_1^{(1)}}=\frac{\partial L}{\partial x_1^{(1)}}\frac{\partial x_1^{(1)}}{\partial o_1^{(1)}}\frac{\partial o_1^{(1)}}{\partial b_1^{(1)}}=\frac{\partial L}{\partial x_1^{(1)}}x_1^{(1)}\left(1-x_1^{(1)}\right) \tag{8}$$

$$\frac{\frac{\partial L}{\partial w_1^{(1)}}}{\frac{\partial L}{\partial b_1^{(1)}}}=x_1^{(0)} \tag{9}$$

从以上各式可以看出, $w_1^{(1)}$ 的梯度与 $b_1^{(1)}$ 的梯度相除便可以推导出输入数据 $x_1^{(0)}$. 采用相同的方法, 便能够逐步推导出全部输入数据. 因此, 在联邦学习过程中, 参与者上传的梯度参数有可能导致其本地隐私数据的泄露. 图 2 为泄露不同比例数据的图片. 从图 2 中可以看出, 攻击者推理出图像中部分信息, 便有可能导致图像中的核心信息泄露.



(a) 数据泄露 100% (a) Leakage 100%
(b) 数据泄露 80% (b) Leakage 80%
(c) 数据泄露 50% (c) Leakage 50%
(d) 数据泄露 30% (d) Leakage 30%

图 2 不同比例数据泄露的图片

Fig. 2 Images of different proportion data leakage

并且, 参数服务器 (云服务器) 拥有全局模型的权重参数 w_{global}, 即本次参与者使用的模型参数. 可以使用所推理出的数据, 按照模型前向传播计算得出 $x_1^{(1)}$ 和 $x_1^{(2)}$, 然后根据式 (12), 便可以求出 y 值, 此时输入数据与其相应的目标输出都泄露给参数服务器.

$$x_1^{(1)}=sigmoid\left(\sum_{i=1}^{N}\left(x_i^{(0)}\cdot w_i^{(0)}\right)+b_1^{(1)}\right) \tag{10}$$

$$x_1^{(2)}=f\left(\boldsymbol{x}^{(0)},\, w_{\text{global}}\right) \tag{11}$$

$$\frac{\partial L}{\partial w_1^{(1)}}=\frac{\partial L}{\partial x_1^{(2)}}\frac{\partial x_1^{(2)}}{\partial o_1^{(2)}}\frac{\partial o_1^{(2)}}{\partial w_1^{(1)}}= \left(x_1^{(2)}-y\right)x_1^{(2)}\left(1-x_1^{(2)}\right)x_1^{(1)} \tag{12}$$

2) 情况 2 (考虑 Dropout)

在深度神经网络中的某一层使用 Dropout, 则该层的神经元会被随机“冻结”, 从而在训练过程中网络模型结构会随机发生变化[32]. 该方法已经广泛用于提高深度学习模型的泛化能力, 防止模型过拟合.

$$x_i^{(l)}=f(\boldsymbol{w}_i^{(l)}\times\boldsymbol{y}^{(l-1)}+b_i^{(l)}) \tag{13}$$

$$\tilde{\boldsymbol{x}}^{(l)}=\boldsymbol{r}^{(l)}\times\boldsymbol{y}^{(l)},\quad r_j^{(l)}\sim\text{Bernoulli}(p) \tag{14}$$

当参与者使用 Dropout 时, 被“冻结”的神经元的输出值为零, 因此不参与神经网络的反向传播计算, 即与这些神经元相连权重的梯度参数为零, 一定程度上能够起到保护参与者本地数据隐私的作用[33]. 然而, 深度学习模型有着较多的神经网络节点, 服务器可以通过有数值的梯度参数进行暴力破解, 从而推理出参与者的隐私信息. 例如, 可以通过任意一个未冻结的节点 j, 根据式 (15) ~ (17) 计算, 获得输入数据 x_i 信息.

$$\frac{\partial L}{\partial w_{i,j}^{(1)}}=\frac{\partial L}{\partial x_j^{(1)}}\frac{\partial x_j^{(1)}}{\partial o_j^{(1)}}\frac{\partial o_j^{(1)}}{\partial w_{i,j}^{(1)}}=\frac{\partial L}{\partial x_j^{(1)}}x_j^{(1)}\left(1-x_j^{(1)}\right)x_i^{(0)} \tag{15}$$

$$\frac{\partial L}{\partial b_{1,j}^{(1)}}=\frac{\partial L}{\partial x_j^{(1)}}\frac{\partial x_j^{(1)}}{\partial o_j^{(1)}}\frac{\partial o_j^{(1)}}{\partial b_{1,j}^{(1)}}=\frac{\partial L}{\partial x_j^{(1)}}x_j^{(1)}\left(1-x_j^{(1)}\right) \tag{16}$$

$$\frac{\frac{\partial L}{\partial w_{i,j}^{(1)}}}{\frac{\partial L}{\partial b_{1,j}^{(1)}}}=x_i^{(0)} \tag{17}$$

2.2 性能损失

联邦学习技术在工程应用时, 相比于传统集中式学习会出现一定的精度损失[15, 34]. 本节以采用随机梯度下降方法更新的深度学习模型为例, 对联邦学习中的模型精度损失进行数学化描述. 在机器学习领域中, 通常假设训练样本与测试样本具有相同的数据分布[15], 即 $\hat{p}_{\text{train}}=\hat{p}_{\text{test}}$. 本文假设集中式学习模型和联邦学习模型使用相同测试集、模型初始参数和学习率, 即 $\eta_{\text{CL}}=\eta_{\text{FL}(k)}$, $w_0^{\text{CL}}=w_0^{\text{FL(global)}}=w_0^{\text{FL}(k)}$.

采用随机梯度下降法的集中式学习的优化目标为

$$\min_{w^{\text{CL}}}\ \text{E}_{(x,\,y)\sim\hat{p}\text{train}}L\left[f(x;w^{\text{CL}}),\,y\right] \tag{18}$$

式中, L 为损失函数, $\hat{p}_{\text{train}}$ 为训练数据的分布, $(x,\,y)$ 为集中式学习方法用的训练数据样本及其相应的标签, w^{CL} 为集中式学习方法的模型参数.

集中式学习模型的参数更新式为

$$w_{e+1}^{\mathrm{CL}}=w_e^{\mathrm{CL}}-\frac{\eta}{n}\left(\sum_{i=1}^{n}\nabla_{w_e^{\mathrm{CL}}}L\left(f\left(x^{(i)};w_e^{\mathrm{CL}}\right),y^{(i)}\right)\right) \tag{19}$$

式中, w_e^{CL} 为第 e 次迭代的模型参数, $x^{(i)}$ 为训练集中的样本, $y^{(i)}$ 为其相应的目标输出.

采用随机梯度下降法的联邦学习优化目标为

$$\min_{w^{\mathrm{FL}(k)}}\mathrm{E}_{(x,\,y)\sim\hat{p}_{\mathrm{train}}^{(k)}}L\left[f(x;w^{\mathrm{FL}(k)}),\,y\right] \tag{20}$$

式中, $\hat{p}_{\mathrm{train}}^{(k)}$ 为参与者 k 的训练数据的分布, $w^{\mathrm{FL}(k)}$ 为参与者 k 的模型参数.

联邦学习中参与者的本地模型权重参数更新式为

$$w_{e(r,\,t+1)}^{\mathrm{FL}(k)}=w_{e(r,\,t)}^{\mathrm{FL}(k)}-\frac{\eta}{n_k}\times\left(\sum_{i=1}^{n_k}\nabla_{w_{e(r,\,t)}^{\mathrm{FL}(k)}}L\left(f\left(x^{(i)};w_{e(r,\,t)}^{\mathrm{FL}(k)}\right),y^{(i)}\right)\right) \tag{21}$$

式中, $w_{e(r,\,t+1)}^{\mathrm{FL}(k)}$ 为参与者 k 在 e 次迭代中的第 $t+1$ 次本地训练的模型权重, n_k 为参与者 k 的样本个数.

当本地训练次数达到第 τ 次, 即满足全局模型更新条件时, 参与者上传更新后的本地模型参数用于参数服务器更新全局模型参数. 在工程应用中, 参与者通常拥有不同数量的样本, 因此联邦学习的全局模型参数采用加权更新, 即

$$\begin{aligned}w_{e(r+1)}^{\mathrm{FL(global)}}&=w_{e(r)}^{\mathrm{FL(global)}}-\\&\frac{\eta}{n}\sum_{k=1}^{N}\left(\sum_{i=1}^{n_k}\nabla_{w_{e(r,\,\tau)}^{\mathrm{FL}(k)}}L\left(f\left(x^{(i)};w_{e(r,\,\tau)}^{\mathrm{FL}(k)}\right),y^{(i)}\right)\right)=\\&\sum_{k=1}^{N}\frac{n_k}{n}w_{e(r,\,\tau)}^{\mathrm{FL}(k)}\end{aligned} \tag{22}$$

分析以上公式, 引起联邦学习模型性能损失的主要原因总结如下:

1) 参与者的数据分布不一致: 根据研究[15], 并结合式 (17) 和式 (18) 可以得出, 当联邦学习所有参与者的训练集分布与集中式学习的训练集分布相同时, 即 $\hat{p}_{\mathrm{train}}^{(k)}=\hat{p}_{\mathrm{train}}=\hat{p}_{\mathrm{test}}$, 理论上联邦学习模型与集中式学习模型可以获得相同的性能, 然而, 实际上参与者的训练集分布往往是不相同的, 即 $\hat{p}_{\mathrm{train}}^{(k)}\neq\hat{p}_{\mathrm{train}}\neq\hat{p}_{\mathrm{test}}$, 因此联邦学习相比于集中式学习模型会出现性能损失.

2) 参与者拥有的数据量不同: 在式 (20) 中, 参与者使用所拥有的全部样本进行训练. 在深度学习模型训练时, 为加快模型收敛速度, 降低内存占用, 通常采用 Mini-batch 的方法训练模型. 例如, 在文献 [24, 33, 35] 提出的联邦学习算法中, 参与者使用 Mini-batch 的方式进行本地模型训练. 然而, 当参与者采用相同的 Mini-batch 设定值进行训练时, 由于参与者拥有的数据量不同, 则不同参与者的样本用于模型训练的概率也不相同. 此时, 数据量分布不均可能会对全局模型的性能造成影响, 并随着迭代次数增加而逐步累积.

3) 全局模型更新间隔不同: 在式 (21) 和式 (22) 中, 当参与者本地训练次数达到第 τ 次时, 参数服务器再进行全局模型参数的更新. 文献 [16] 指出, 当 $\tau=1$ 时, 联邦学习理论上能够与集中式学习有相同的性能. 在本地模型属于凸模型的假设前提下, 文献 [17] 指出, 全局模型更新间隔会影响到联邦学习模型的性能, 如式 (23) 所示:

$$L(w^{\mathrm{FL}})-L(w^{*})\leq\frac{1}{2\eta\varphi T}+\sqrt{\frac{1}{4\eta^2\varphi^2T^2}+\frac{\rho h(\tau)}{\eta\varphi\tau}}+\rho h(\tau) \tag{23}$$

式中, w^{*} 为最优模型参数值, T 为参与者本地模型训练总次数, ρ 为 Lipschitz 参数.

然而, 为降低计算和通信成本, 很多研究[16, 18, 33]设定的模型聚合间隔 τ 大于 1. 在数据非独立同分布的场景下, 过大的全局模型更新间隔, 会造成联邦学习模型性能严重下降, 甚至导致模型不收敛.

3　支持隐私保护的自适应联邦学习算法

3.1　安全模型与目标

本研究假设各参与者和参数服务器为“诚实且好奇”的半可信实体. 在联邦学习研究领域中, 半诚实模型是一种常见的假设模型[24−25, 27]. 根据该假设模型, 本文中参与者和参数服务器都会遵守所设定的协议, 但是在训练期间都想通过中间数据推理获得其他参与者的数据隐私信息. 本文的研究目标是, 在联邦学习训练过程中, 参数服务器不能获得参与者的敏感信息 (如训练样本及模型参数), 同时参与者也不能获得其他参与者的敏感信息.

3.2　联邦学习过程

如图 3 所示, 本文提出的联邦学习系统包含 N 个参与者和 1 个参数服务器, 并且每个参与者都拥有本地数据库. 参与者上传本地模型的权重参数, 由第三方参数服务器实现全局模型参数更新等功能. 所提出的联邦学习系统, 主要包含 3 个阶段: 系统初始化、系统训练和模型部署.

3.2.1　系统初始化

采用文献 [18] 提出的联邦学习信息交互方案,

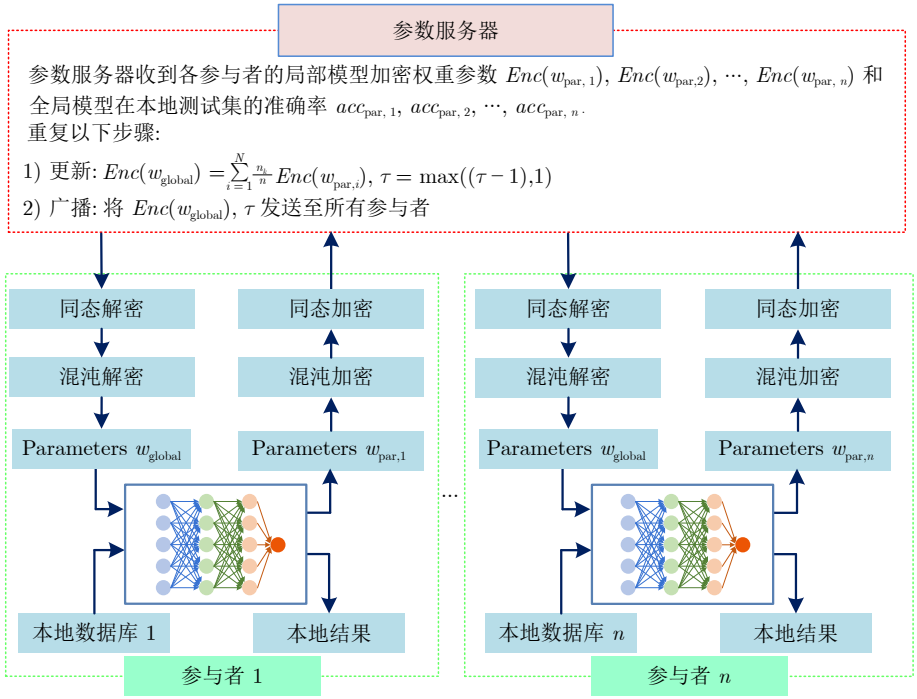


图 3 本文所提出的联邦学习系统结构图

Fig. 3 The structure diagram of the proposed federated learning system

各参与者与参数服务器建立不同的 TLS (Transport layer security)/SSL (Secure sockets layer) 通道, 以保证通信安全. 参与者们采用 AES (Advanced encryption standard) 加密的方式, 互相沟通确定同态加密算法的公钥 PK 和私钥 SK、混沌系统的初始状态参数、深度学习模型的超参数. 特别地, 各参与者对同态加密算法的公钥、私钥和混沌系统的初始状态参数保密, 不会泄露给任何非参与的实体, 例如参数服务器和系统外用户. 首先, 各参与者生成本地的深度学习模型, 并将初始的模型参数进行加密上传至参数服务器. 然后, 参数服务器聚合所有参与者上传的模型参数密文, 生成全局模型初始参数密文并广播至各个参与者. 最后, 参与者对全局模型参数密文进行解密并加载至本地模型中.

3.2.2 系统训练

1) 参与者

如图 4 所示, 参与者首先从参数服务器下载全局模型参数密文和全局模型更新间隔设定值. 各参与者根据 CKKS 加密方案, 使用私钥 SK 对全局模型权重参数密文进行解密, 完成第 1 阶段解密. 然后使用混沌系统生成的随机数, 对上述数据进行第 2 阶段的混沌解密, 从而获得全局模型的权重参数 w_{global}, 并将 w_{global} 载入本地模型中. 接下来, 使用本地测试集对载入 w_{global} 的本地模型进行准确率测试, 得到本地测试准确率 $acc_{\text{par},i}$. 根据本文所提出的自适应模型聚合方案 (具体见第 3.3 节) 设置 Mini-batch 值, 并使用本地数据集对模型进行训练. 当训练 Mini-batch 的次数达到聚合方案设定的全局模型更新频率, 即满足全局模型更新间隔 τ 时, 参与者使用混沌系统对本地模型的权重参数进行置乱加密, 从而完成第 1 阶段的混沌加密. 特别地, 各参与者使用相同的初始状态参数, 以保证参数服务器运算的同态性. 然后, 使用 CKKS 加密方案对模型参数进行第 2 阶段的同态加密. 最后, 将加密的本地模型参数和本地测试准确率 $acc_{\text{par},i}$ 上传至参数服务器, 进行下一次迭代学习或结束.

2) 参数服务器

参数服务器根据自适应模型聚合方法确定全局模型更新间隔 τ. 参数服务器将收到参与者密文 $Enc(w_{\text{par},i})$, 使用式 (24) 对全局模型参数加权更新

$$Enc(w_{\text{global}}) = \sum_{i=1}^{N} \frac{n_k}{n} Enc(w_{\text{par},i}) \tag{24}$$

接下来, 服务器根据本文所提出的自适应模型聚合方案 (具体见第 3.3 节) 对联邦学习系统中的全局模型聚合间隔 τ 进行调整. 最后, 参数服务器将更新后的模型参数密文和全局模型聚合间隔 τ 广播至各参与者.

3.2.3 模型部署

当联邦学习过程完成后, 所有参与者断开与参

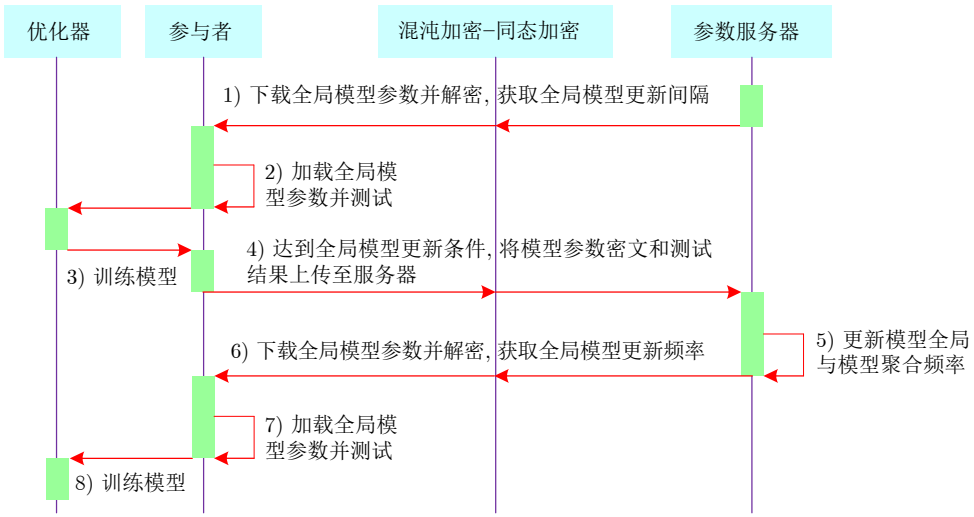


图 4　联邦学习训练过程交互图

Fig. 4　Interaction diagram of the federated learning system

数服务器的通信连接, 不再与参数服务器进行数据交互, 同时停止本地模型训练. 将训练好的全局模型参数加载至本地模型, 用于本地任务的使用.

3.3　自适应模型聚合方案

本文提出的联邦学习模型自适应模型聚合方案包括两个部分: 参与者 Mini-batch 值设定和自适应调整全局模型更新间隔算法.

1) 参与者 Mini-batch 值设定

假设联邦学习使用的数据分布为非独立同分布, 即各参与者拥有不同类别的数据, 并且各参与者聚合数据分布与集中式学习采用的训练集和测试集的数据分布相同

$$aggregate\left(\hat{p}_{\text{train}}^{(1)}, \cdots, \hat{p}_{\text{train}}^{(N)}\right) = \hat{p}_{\text{train}} = \hat{p}_{\text{test}} \tag{25}$$

假设参与者采用 Mini-batch 方法从其本地数据库随机抽取数据时, 每个 Mini-batch 的数据分布都与参与者的训练数据分布相同, 即 $\hat{p}_{\text{minibatch}}^{(k)} = \hat{p}_{\text{train}}^{(k)}$. 各参与者抽取 Mini-batch 值的确定采用以下计算:

$$n_{\text{minibatch}}^{(k)} = n_{\text{minibatch}}^{\text{CL}} \cdot \left(\frac{n_k}{\sum_{k=1}^{N} n_k} \right) \tag{26}$$

式中, $n_{\text{minibatch}}^{(k)}$ 为参与者 k 设定的 Mini-batch 值, $n_{\text{minibatch}}^{\text{CL}}$ 为集中式学习设定的 Mini-batch 值, n_k 为参与者 k 的样本数量. 结合式 (25) 和式 (26), 可以得到

$$aggregate\left(\hat{p}_{\text{minibatch}}^{(1)}, \cdots, \hat{p}_{\text{minibatch}}^{(N)}\right) = \hat{p}_{\text{train}} = \hat{p}_{\text{test}} \tag{27}$$

$$n_{\text{minibatch}}^{\text{CL}} = \sum_{k=1}^{N} n_{\text{minibatch}}^{(k)} \tag{28}$$

结合式 (21), 可以得到

$$\begin{aligned} w_{e(r+1)}^{\text{FL(global)}} = & \sum_{k=1}^{K} \frac{n_{\text{minibatch}}^{(k)}}{n} w_{e(r,0)}^{\text{FL}(k)} - \\ & \frac{\eta}{n} \sum_{k=1}^{N} \sum_{i=1}^{n_{\text{minibatch}}^{(k)}} \nabla_{w_{e(r,\tau)}^{\text{FL}(k)}} L\left(f\left(x^{(i)}; w_{e(r,\tau)}^{\text{FL}(k)}\right), y^{(i)}\right) = \\ & w_{e(r)}^{\text{FL(global)}} - \\ & \frac{\eta}{n} \sum_{k=1}^{N} \sum_{i=1}^{n_{\text{minibatch}}^{(k)}} \nabla_{w_{e(r,\tau)}^{\text{FL}(k)}} L\left(f\left(x^{(i)}; w_{e(r,\tau)}^{\text{FL}(k)}\right), y^{(i)}\right) = \\ & w_{e(r+1)}^{\text{CL}}, \quad \tau = 1 \end{aligned} \tag{29}$$

式中, $w_{e(r+1)}^{\text{FL(global)}}$ 为联邦学习在第 e 次迭代的第 $r+1$ 轮 Mini-batch 的模型权重参数. 通过以上数学公式推导可以看出, 在 $\tau = 1$ 时采用本文提出的 Mini-batch 自适应方法, 理论上能够获得与集中式学习模型相同的性能.

2) 自适应调整全局模型更新间隔算法

在非独立同分布数据的情况下, 采用较大的全局模型更新间隔 τ [33] 会导致全局模型的精度降低, 而采用较小的更新间隔 τ 则可能导致通信成本大大提高, 降低全局模型训练效率.

因此, 本文参考深度学习中的自适应调整学习率的方法, 提出一种自适应调整全局模型更新间隔方法. 在模型训练过程中, 参数服务器首先聚合参与者们上传的本地数据测试准确率, 其计算式为

$$acc = \sum_{i=1}^{N} \frac{n_k}{n} acc_{\text{par},\,i} \tag{30}$$

在训练过程中, 若连续 Φ 次没达到历史训练过程的最高精度, 则按照式 (31) 调整全局模型更新间隔 τ:

$$\tau = \max((\tau - 1),\, 1) \tag{31}$$

本文所提出的自适应调整全局模型更新间隔算法, 能够根据参与者训练的反馈结果, 实时调整全局模型更新间隔, 从而提高联邦学习训练效率, 同时保证模型精度.

4 自适应联邦学习算法分析

4.1 安全性分析

定义 1 (CPA 安全). 如果对于所有概率多项式时间 (Probabilistic polynomial-time, PPT) 敌手 $\mathcal{A}$, 存在一个可以忽略的函数 $negl$, 使得

$$\Pr\left[Priv\left(K^{\text{cpa}}_{\mathcal{A},\,\Pi}(n)\right) = 1\right] \leq \frac{1}{2} + negl(n) \tag{32}$$

则称密钥加密方案 $\Pi = (Gen,\ Enc,\ Dec)$ 是符合 CPA 安全的加密方案. 其中, 概率的来源是敌手 $\mathcal{A}$ 的随机性和实验的随机性 (生成密钥 PK、随机比特以及在加密过程中出现的各种随机性).

通过以上定义可以得出:

1) 所有满足 CPA 安全的加密方案同样也是满足窃听者存在情况下的安全加密方案;

2) 任何确定性的加密方案都不满足 CPA 安全, 满足 CPA 安全的加密方案一定是概率加密.

定理 1. 在本文提出的联邦学习方案中, 如果同态加密 CKKS 方案是 CPA 安全的, 同时所有参与者和参数服务器/外部攻击者之间没有串谋, 则该方案能够保护参与者的数据隐私信息.

证明. 假设存在一个敌手 $\mathcal{A}$, 窃取了所有加密的模型权重参数. 由于敌手 $\mathcal{A}$ 不知道 CKKS 方案的设定值 λ, 从而 $\mathcal{A}$ 不能生成密钥 SK. 根据本文的安全假设, 参与者不会与服务器和系统外部成员串谋, 从而密钥 SK 不会泄露给参与者外的其他实体, 所以敌手 $\mathcal{A}$ 不会获得密钥 SK. 因此, 敌手 $\mathcal{A}$ 无法对模型参数的权重密文进行解密, 从而获取模型权重参数的真实值. 同时, 模型的权重参数以密文的形式存储在服务器上, 只要参数服务器不与其他参与者串谋, 则参与者获取不到其他参与者上传的模型权重参数. 同时, 参与者通过不同的安全通信通道与参数服务器进行信息传输, 从而防止传输的信息被窃取. 因此, 该联邦学习方案能够保护参与者的隐私信息不被泄露. □

为进一步提升联邦学习隐私保护水平, 本文参考文献 [31] 中使用模型参数置乱的方式保护参与者的数据隐私信息. 基于此, 将文献 [28] 中提出的基于混沌系统的置乱加密算法, 引入到本文提出的联邦学习系统中. 如图 5 所示, 通过加密的共享模型参数 (不使用同态加密算法) 推断出的信息等同于该算法对输入图像进行加密. 因此, 我们可以利用图像加密领域常用的信息熵 (Information entropy, IE) 来对推断的数据进行分析. 从实验图中可以看出, 从混沌加密的模型参数中推理得到的图片信息熵接近于理想值 8. 因此, 采用混沌加密技术能够进一步提升联邦学习的信息安全水平.

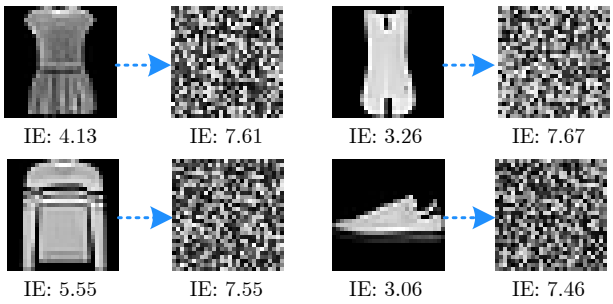


图 5 加密参数的推理数据图

Fig. 5 Inferring data of the encrypted parameters

4.2 算法性能分析

相比于传统集中式训练方法, 本文提出的联邦学习算法包含的额外时间开销主要为模型的权重参数加密/解密、参与者上传/下载模型参数和参数服务器更新全局模型参数. 本文采用混沌系统生成伪随机数对模型参数进行第 1 阶段的加密, 其时间消耗的主要部分为伪随机数的生成. 使用 CKKS 的操作主要为加密、解密以及密文操作. 如表 1 所示, 在时间开销上采用 CKKS 加密的联邦学习算法明显优于基于 Paillier 加密的机器学习算法[18−19, 23], 并且混沌系统生成伪随机数消耗的时间也在接受范围内.

表 1 加密/解密算法的执行时间

Table 1 Execution time of the encryption/decryption operations

操作类型	500 个参数	2000 个参数	54000 个参数
随机数生成	12.05 ms	25.50 ms	0.40 s
CKKS 加密	9.37 ms	9.68 ms	0.54 s
CKKS 解密	1.56 ms	17.18 ms	0.03 s
CKKS 密文加法	0.15 ms	0.15 ms	0.02 s
Paillier 加密	3.82 s	14.61 s	410.32 s
Paillier 解密	1.06 s	4.22 s	115.92 s
Paillier 密文加法	7.87 ms	30.03 ms	0.87 s

假设每个参与者都有 20 000 个样本, Mini-batch 为 128, 则联邦学习系统采用不同数值的全局模型更新间隔[16, 18, 33] 的模型参数加密/解密的次数如表 2 所示. 从表 2 中可以看出, 减小全局模型更新间隔 τ, 即提高全局模型更新频率, 会导致联邦学习加密解密运算次数大幅增加. 因此, 本文采用动态全局模型更新间隔方法, 在保证模型精度的前提下, 提高系统训练效率.

表 2　加密/解密算法的执行次数

Table 2　Execution numbers of the encryption/decryption operations

模型更新间隔	50 次	80 次	100 次
$\tau=1$ [18]	7 800	12 500	15 600
$\tau=4$ [16]	1 550	2 480	3 100
$\tau=15$ [33]	500	800	1 000

4.3　功能对比分析

表 3 将本文提出的联邦学习算法 APFL (Adaptive privacy-preserving federated learning) 与近期研究提出的联邦学习算法进行功能性的对比, 其中包括 PFL (Privacy-preserving federated learning)[18]、AFL (Adaptive federated learning)[17] 和 MFL (Momentum federated learning)[16]. PFL 算法虽然能够对联邦学习系统中的参与者进行数据隐私保护, 但是没有对联邦学习训练过程进行优化或提高. AFL 和 MFL 分别使用自适应聚合频率算法和动量梯度下降算法以提高联邦学习的训练效率. 然而, 这两种方法都没有考虑对参与者的数据隐私进行保护. 此外, 以上三种联邦学习算法都不涉及对参与者的 Mini-batch 值进行设定. 本文提出的 APFL 算法利用同态加密技术和混沌加密技术对参与者的数据隐私进行保护, 同时提出自适应模型聚合方案和采用动量梯度下降法提高联邦学习训练效率, 降低计算资源和通信资源的消耗.

表 3　不同联邦学习方案的功能分析

Table 3　The functionality analysis of the different FLs

功能	PFL	AFL	MFL	APFL
隐私保护	√	×	×	√
自适应调整 τ	×	√	×	√
Mini-batch 设定	×	×	×	√
动量项加速	×	×	√	√

5　实验与分析

5.1　实验环境及实现

实验环境为 Windows10, MATLAB2018b, Python 3.6, Pytorch1.5 和 CUDA10.1 用于搭建深度神经网络模型, Python 通过 Matlab-Python 接口调用 MATLAB 中的混沌系统程序生成伪随机数, 使用开源的 CKKS 库对模型的权重参数进行加密与解密操作. Fashion-MNIST (F-MNIST) 和 CIFAR10 数据集用于验证本文所提出的联邦学习系统的有效性.

上述两个数据集各包含有 10 个不同类别的图片, 训练集和测试集分别有 60 000 张图片和 10 000 张图片. 在实际场景中, 各参与者的数据库拥有的数据类别往往是不相同的. 因此, 本文将数据集切分为非独立同分布数据集, 用于对所提出的联邦学习算法进行评估. 具体方式为: 训练集中类别编号为 0 ~ 3 的图像划分为数据集 N-train1, 编号为 4 ~ 6 的图像划分为数据集 N-train2, 编号为 7 ~ 9 的图像划分为数据集 N-train3. 将分割好的数据集 N-train1, N-train2 和 N-train3 分别用于参与者 1, 2 和 3 的本地模型训练. 设置训练迭代次数 Epoch 为 50, 优化算法为小批量梯度下降 (Mini-batch gradient descent, MGD) 算法 (学习率 $\eta=0.1$, 动量 $\gamma=0.5$), 对应集中式学习 CL 的 Mini-batch 值为 512. 自适应全局模型聚合间隔算法的初始全局模型聚合间隔 τ 设为 15, 连续未达到精度要求 Φ 值设为 5.

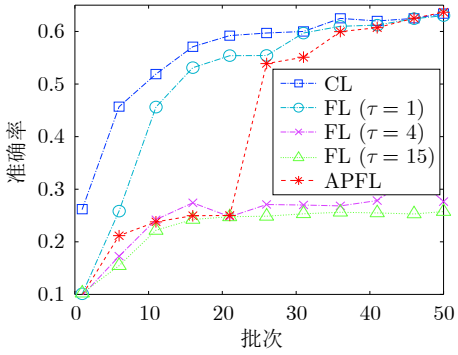
采用准确率 (Precision)、查全率 (Recall) 和综合评价指标 (F1-score) 进行算法评估[36]. 采用文献 [19] 中提出的方法, 评估联邦学习模型与集中式学习模型之间的偏差, 即

$$Dev_{\text{avg}}=\frac{1}{cn}\sum_{i=1}^{cn}|acc_{\text{FL},i}-acc_{\text{CL},i}| \tag{33}$$

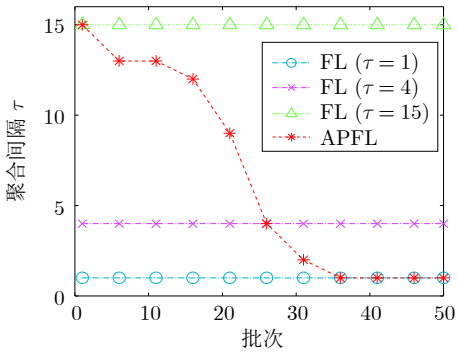
式中, cn 为数据集中类别的个数, $acc_{\text{FL},i}$ 为联邦学习模型在第 i 类数据识别的准确率, $acc_{\text{CL},i}$ 为集中式学习模型在第 i 类数据识别的准确率.

5.2　性能对比实验

根据文献 [16, 18, 33], 本节分别设置模型更新间隔为 $\tau=1$, 4 和 15, 与采用本文提出的自适应联邦学习算法进行对比. 自适应联邦学习算法首先根据所提出的自适应模型聚合方案对各个参与者的 Mini-batch 值进行设定, 联邦学习中参与者 1 号的 Mini-batch 设为 $\text{int}(512\times(4/10))=204$, 而参与者 2 号和 3 号的 Mini-batch 设为 $\text{int}\,(512\times(3/10))=153$. 图 6 和图 7 分别为各联邦学习方案在 CIFAR10 和 F-MNIST 数据集上的实验曲线. 从准确率曲线可以看出, 所提出方法在训练过程前期模型准确率上升速度低于 FL ($\tau=1$, 4) 和 CL, 因为此时 APFL 所设定的 τ 大于 4. 随着训练过程的进行,



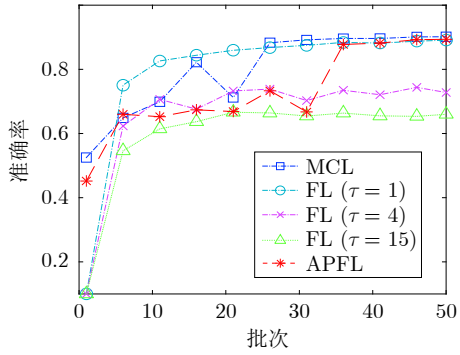
(a) 准确度曲线
(a) Accuracy curves



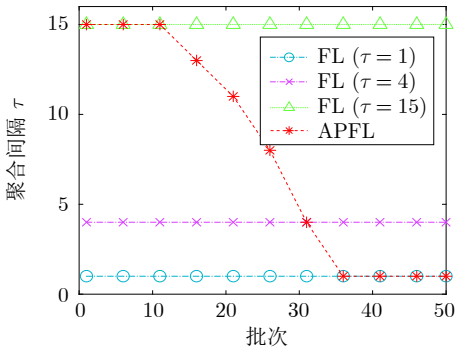
(b) 更新频率曲线
(b) Aggregation frequency curves

图 6 在 CIFAR10 上不同联邦学习模型的实验曲线

Fig. 6 Experiment curves of the different federated learning models on CIFAR10



(a) 准确度曲线
(a) Accuracy curves



(b) 更新频率曲线
(b) Aggregation frequency curves

图 7 在 F-MNIST 上不同联邦学习模型的实验曲线

Fig. 7 Experiment curves of the different federated learning models on F-MNIST

APFL 逐渐调低全局模型聚合间隔 τ, 即提升全局模型更新频率. 表 4 和表 5 分别为不同联邦学习模型在 CIFAR10 和 F-MNIST 数据集上的分类结果. 从表中可以看出, 在 CIFAR10 和 F-MNIST 数据集上, APFL 的通信次数相比于 FL ($\tau = 1$) 分别降低了 31.43% 和 55.44%. 同时, 从实验结果可以看出, 本文所提出的自适应调整模型聚合间隔方法, 能够根据任务难度进行调整. 在较为简单的 F-MNIST 数据集, 全局模型聚合间隔降低的速率低于 CIFAR10 数据集. 从本节的实验结果可以看出, 所提出的自适应模型聚合方案能够在保证模型精度的前提下, 降低模型聚合次数, 从而降低计算和通信成本, 进而提高训练效率.

表 4 CIFAR10 上不同联邦学习模型的分类结果 (%)

Table 4 Classification results of the different federated learning models on CIFAR10 (%)

方法	准确率	精准率	召回率	Dev_{avg}	聚合次数
CL	63.36	63.92	63.29	—	—
FL ($\tau = 15$) [33]	25.76	9.34	25.87	49.91	250
FL ($\tau = 4$) [16]	27.64	50.14	27.76	45.04	1 100
FL ($\tau = 1$) [18]	61.78	62.76	61.77	1.91	4 400
APFL	**63.66**	**63.49**	**63.64**	**2.02**	**2 758**

表 5 F-MNIST 上不同联邦学习模型的分类结果 (%)

Table 5 Classification results of the different federated learning models on F-MNIST (%)

方法	准确率	精准率	召回率	Dev_{avg}	聚合次数
CL	90.15	90.07	90.15	—	—
FL ($\tau = 15$) [33]	65.99	62.18	65.99	31.43	350
FL ($\tau = 4$) [16]	72.77	65.24	72.77	23.16	1 350
FL ($\tau = 1$) [18]	89.10	89.25	89.10	0.88	5 250
APFL	**89.36**	**89.30**	**89.36**	**0.87**	**2 339**

5.3 自适应模型聚合方案作用分析

本文提出的自适应模型聚合方案主要包含两个部分: Mini-batch 设定和自适应调整全局模型聚合间隔算法. 为评估这两个部分的作用, 本节采用消融实验对其进行分析.

5.3.1 Mini-batch 设定消融实验

第 3.3 节提出了不同参与者 Mini-batch 设定

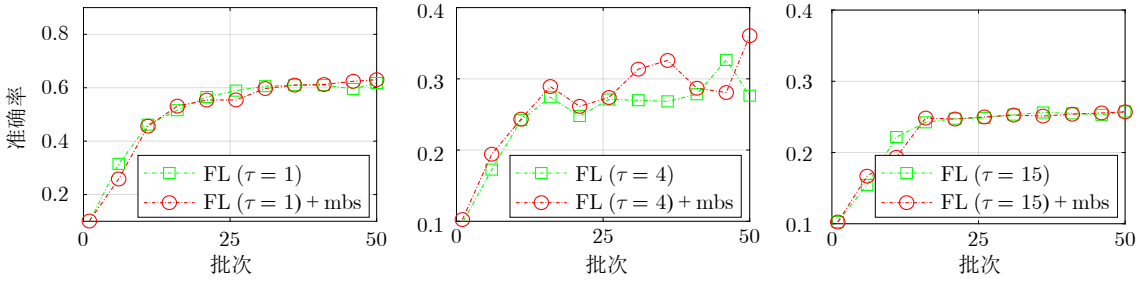
方案, 该方案根据参与者所拥有的数据量大小, 对参与者的 Mini-batch 值进行设定. Mini-batch 设定的消融实验结果如图 8、表 6 和表 7 所示. 由实验结果可以看出, 联邦学习模型使用了 Mini-batch 设定算法后, 能够一定程度上提升模型的性能. 并且, 在较低的全局模型更新间隔 $\tau = 4$ 或较为简单的数据集 F-MNIST 时, 采用 Mini-batch 设定算法对联邦学习的性能提升效果较为明显.

5.3.2　自适应调整全局模型更新间隔算法消融实验

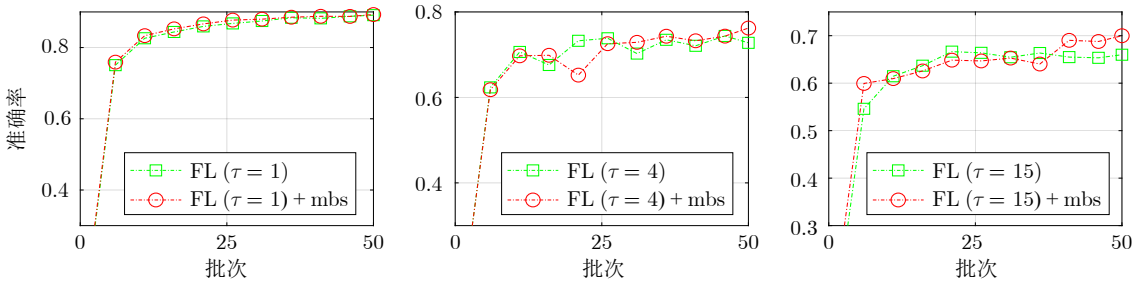
第 3.3 节提出了自适应全局模型更新间隔方法, 该方法能够根据参与者上传的训练信息, 对全局模型更新间隔进行调整. 自适应全局模型更新间隔方法的消融实验结果如图 9、图 10、表 8 和表 9 所示. 由实验结果可以看出, 采用了自适应调整全局模型更新间隔算法的联邦学习模型性能与 FL (τ =1) 模型接近相同, 同时减小了模型聚合次数, 进而提高了联邦学习训练效率. 值得注意的是, 相比于 APFL 算法, 没有采用 Mini-batch 设定算法 (APFL (no mbs)) 的模型性能发生了下降, 与 CL 模型的精度偏差有所增加.

6　结束语

本文首先以神经网络模型为例, 说明梯度参数是如何泄露本地数据的, 并通过数学公式推导分析了联邦学习性能损失的原因. 针对联邦学习性能损



(a) 在CIFAR10 数据集上的精度曲线
(a) Accuracy w.r.t CIFAR10



(b) 在 F-MNIST 数据集上的精度曲线
(b) Accuracy w.r.t F-MNIST

图 8　在 CIFAR10 和 F-MNIST 数据集的 Mini-batch 设定消融实验曲线

Fig. 8　Experiment curves of the Mini-batch size setting on CIFAR10 and F-MNIST

表 6　CIFAR10 下的 Mini-batch 设定消融实验结果 (%)

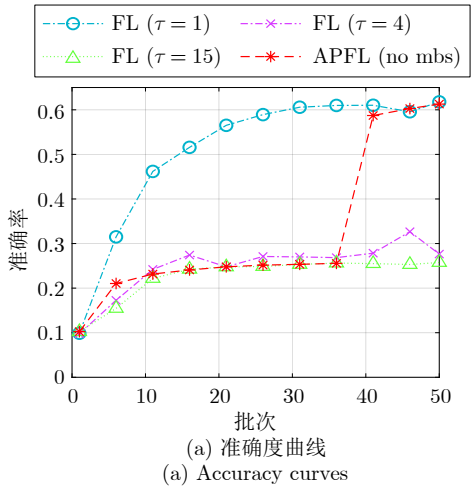
Table 6　Ablation experiment results of the Mini-batch size setting on CIFAR10 (%)

方法	Accuracy	Precision	Recall	Dev_{avg}
CL	63.36	63.92	63.29	—
FL ($\tau = 15$) [33]	25.76	9.34	25.87	49.91
FL ($\tau = 15$) + mbs	25.70	9.14	25.78	50.07
FL ($\tau = 4$)	27.64	50.14	27.76	45.04
FL ($\tau = 4$) + mbs	**63.66**	60.93	36.06	32.90
FL ($\tau = 1$) [18]	61.78	62.76	61.77	1.91
FL ($\tau = 1$) + mbs	63.02	64.08	62.27	**1.53**

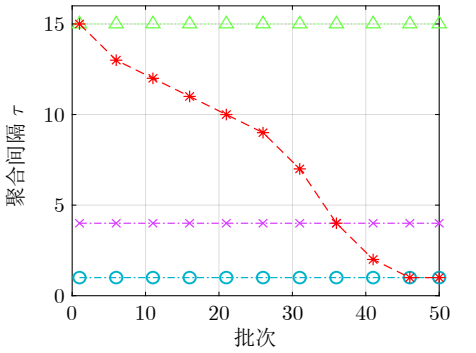
表 7　F-MNIST 下的 Mini-batch 设定消融实验结果 (%)

Table 7　Ablation experiment results of the Mini-batch size setting on F-MNIST (%)

方法	Accuracy	Precision	Recall	Dev_{avg}
CL	90.15	90.07	90.15	—
FL ($\tau = 15$) [33]	65.99	62.18	65.99	31.43
FL ($\tau = 15$) + mbs	69.99	64.29	69.99	26.05
FL ($\tau = 4$)	27.76	50.14	27.76	45.04
FL ($\tau = 4$) + mbs	76.23	84.84	76.23	14.85
FL ($\tau = 1$) [18]	89.10	89.25	89.10	0.88
FL ($\tau = 1$) + mbs	89.27	89.25	89.27	**0.99**



(a) 准确度曲线
(a) Accuracy curves



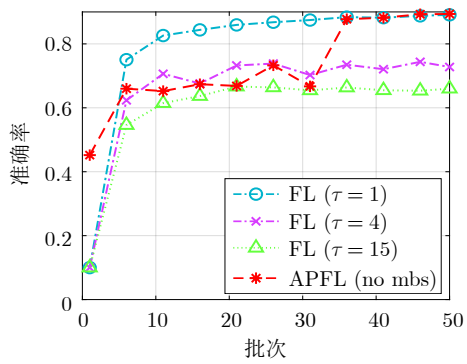
(b) 更新频率曲线
(b) Aggregation frequency curves

图 9 CIFAR10 自适应模型更新间隔消融实验曲线

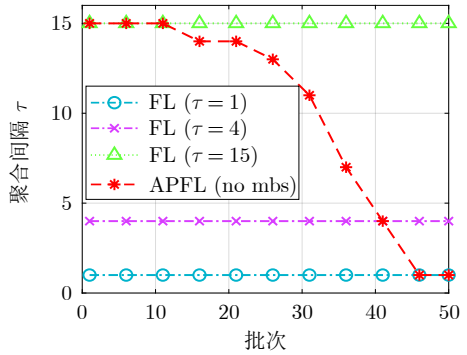
Fig. 9 Experiment curves of the adaptive model aggregation interval on CIFAR10

失问题, 本文提出一种自适应模型聚合方案, 该方案能够自适应调整参与者 Mini-batch 值和全局模型更新间隔. 针对联邦学习隐私泄露的问题, 本文首次将图像加密领域中的混沌加密算法引入联邦学习领域中, 用于构建一种基于混沌系统和同态加密的混合隐私保护方案, 从而进一步提高数据隐私保护水平. 理论分析和实验结果表明, 本文提出的联邦学习算法能够保护参与者的隐私信息, 并在非独立同分布数据的场景下提升训练效率, 降低模型的性能损失, 具备实际工业场景应用的可行性.

本文所提出的联邦学习算法, 没有对低质量数据的参与者进行考虑. 然而, 在真实工业场景中, 可能存在拥有低质量数据的参与者, 进而导致整个联邦学习模型性能下降. 因此, 下一步拟打算在保护参与者隐私前提下, 设计一种低质量数据参与者的识别算法, 并使用更加复杂的数据集对算法进行测试与优化, 从而进一步推动联邦学习在工业领域中的应用.



(a) 准确度曲线
(a) Accuracy curves



(b) 更新频率曲线
(b) Aggregation frequency curves

图 10 F-MNIST 自适应模型更新间隔消融实验曲线

Fig. 10 Experiment curves of the adaptive model aggregation interval on F-MNIST

表 8 CIFAR10 下的自适应更新间隔消融实验结果 (%)

Table 8 Ablation experiment results of the adaptive model aggregation interval on CIFAR10 (%)

方法	Accuracy	Precision	Recall	Dev_{avg}	聚合次数
CL	63.36	63.92	63.29	—	—
FL ($\tau=15$) [33]	25.76	9.34	25.87	49.91	250
FL ($\tau=4$) [16]	27.64	50.14	27.76	45.04	1 100
FL ($\tau=1$) [18]	61.78	62.76	61.77	1.91	4 000
APFL (no mbs)	61.10	62.00	61.36	3.27	1 742

表 9 F-MNIST 下的自适应更新间隔消融实验结果 (%)

Table 9 Ablation experiment results of the adaptive model aggregation interval on F-MNIST (%)

方法	Accuracy	Precision	Recall	Dev_{avg}	聚合次数
CL	90.15	90.07	90.15	—	—
FL ($\tau=15$) [33]	65.99	62.18	65.99	31.43	250
FL ($\tau=4$) [16]	72.77	65.24	72.77	23.16	1 100
FL ($\tau=1$) [18]	89.10	89.25	89.10	0.88	4 400
APFL (no mbs)	89.48	89.42	89.48	0.84	1 336

References

1 Sun Chang-Yin, Mu Chao-Xu. Important scientific problems of multi-agent deep reinforcement learning. *Acta Automatica Sinica*, 2020, **46**(7): 1301−1312
(孙长银, 穆朝絮. 多智能体深度强化学习的若干关键科学问题. 自动化学报, 2020, **46**(7): 1301−1312)

2 Jin Xia-Ting, Wang Yao-Nan, Zhang Hui, Liu Li, Zhong Hang, He Zhen-Dong. DeepRail: Automatic visual detection system for railway surface defect using Bayesian CNN and attention Network. *Acta Automatica Sinica*, 2019, **45**(12): 2312−2327
(金侠挺, 王耀南, 张辉, 刘理, 钟杭, 贺振东. 基于贝叶斯 CNN 和注意力网络的钢轨表面缺陷检测系统. 自动化学报, 2019, **45**(12): 2312−2327)

3 Zhang Z H, Guan C, Liu Z Y. Real-time optimization energy management strategy for fuel cell hybrid ships considering power sources degradation. *IEEE Access*, 2020, **8**: 87046−87059

4 Chen H, Zhang Z H, Guan C, Gao H B. Optimization of sizing and frequency control in battery/supercapacitor hybrid energy storage system for fuel cell ship. *Energy*, 2020, **197**: Article No. 117285

5 Xian Zheng-Zheng, Li Qi-Liang, Huang Xiao-Yu, Lv Wei, Lu Ji-Yuan. Collaborative filtering via SVD++ with differential privacy. *Control and Decision*, 2019, **34**(1): 43−54
(鲜征征, 李启良, 黄晓宇, 吕威, 陆寄远. 基于差分隐私和 SVD++ 的协同过滤算法. 控制与决策, 2019, **34**(1): 43−54)

6 Li J, Kuang X H, Lin S J, Ma X, Tang Y. Privacy preservation for machine learning training and classification based on homomorphic encryption schemes. *Information Sciences*, 2020, **526**: 166−179

7 Gong M G, Pan K, Xie Y, Qin A K, Tang Z D. Preserving differential privacy in deep neural networks with relevance-based adaptive noise imposition. *Neural Networks*, 2020, **125**: 131−141

8 Zhang Chao, Li Qiang, Chen Zi-Hao, Li Zu-Rui, Zhang Zhen. Medical Chain: Alliance medical blockchain system. *Acta Automatica Sinica*, 2019, **45**(8): 1495−1510
(张超, 李强, 陈子豪, 黎祖睿, 张震. Medical Chain: 联盟式医疗区块链系统. 自动化学报, 2019, **45**(8): 1495−1510)

9 Yang Q, Liu Y, Chen T J, Tong Y X. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology*, 2019, **10**(2): Article No. 12

10 Li T, Sahu A K, Talwalkar A, Smith V. Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, 2020, **37**(3): 50−60

11 Zhang W, Li X, Ma H, Luo Z, Li X. Federated learning for machinery fault diagnosis with dynamic validation and self-supervision. *Knowledge-Based Systems*, 2021, **213**: Article No. 106679

12 Sheller M J, Edwards B, Reina G A, Martin J, Pati S, Kotrotsou A, et al. Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data. *Scientific Reports*, 2020, **10**(1): Article No. 12598

13 Kwon D, Jeon J, Park S, Kim J, Cho S. Multiagent DDPG-based deep learning for Smart Ocean federated learning IoT networks. *IEEE Internet of Things Journal*, 2020, **7**(10): 9895−9903

14 Rothchild D, Panda A, Ullah E, Ivkin N, Stoica I, Braverman V, et al. FetchSGD: Communication-efficient federated learning with sketching. In: Proceedings of the 37th International Conference on Machine Learning. Vienna, Austria: JMLR.org, 2020. Article No. 764

15 Duan M M, Liu D, Chen X Z, Liu R P, Tan Y J, Liang L. Self-balancing federated learning with global imbalanced data in mobile systems. *IEEE Transactions on Parallel and Distributed Systems*, 2021, **32**(1): 59−71

16 Liu W, Chen L, Chen Y F, Zhang W Y. Accelerating federated learning via momentum gradient descent. *IEEE Transactions on Parallel and Distributed Systems*, 2020, **31**(8): 1754−1766

17 Wang S Q, Tuor T, Salonidis T, Leung K K, Makaya C, He T, et al. Adaptive federated learning in resource constrained edge computing systems. *IEEE Journal on Selected Areas in Communications*, 2019, **37**(6): 1205−1221

18 Li Q B, Wen Z Y, Wu Z M, Hu S X, Wang N B, Li Y, et al. A survey on federated learning systems: Vision, hype and reality for data privacy and protection. *IEEE Transactions on Knowledge and Data Engineering*, 2023, **35**(4): 3347−3366

19 Zhang Ze-Hui, Fu Yao, Gao Tie-Gang. Research on federated deep neural network model for data privacy preserving. *Acta Automatica Sinica*, 2022, **48**(5): 1273−1284
(张泽辉, 富瑶, 高铁杠. 支持数据隐私保护的联邦深度神经网络模型研究. 自动化学报, 2022, **48**(5): 1273−1284)

20 Lv L J, Li Y T, Nandakumar K, Yu J S, Ma X J. How to democratise and protect AI: Fair and differentially private decentralised deep learning. *IEEE Transactions on Dependable and Secure Computing*, 2022, **19**(2): 1003−1017

21 Wang Y F, Gu M, Ma J H, Jin Q. DNN-DP: Differential privacy enabled deep neural network learning framework for sensitive crowdsourcing data. *IEEE Transactions on Computational Social Systems*, 2020, **7**(1): 215−224

22 Carpov S, Gama N, Georgieva M, Troncoso-Pastoriza J R. Privacy-preserving semi-parallel logistic regression training with fully homomorphic encryption. *BMC Medical Genomics*, 2019, **13**(7): Article No. 88

23 Song Lei, Ma Chun-Guang, Duan Guang-Han, Yuan Qi. Privacy-preserving logistic regression on vertically partitioned data. *Journal of Computer Research and Development*, 2019, **56**(10): 2243−2249
(宋蕾, 马春光, 段广晗, 袁琪. 基于数据纵向分布的隐私保护逻辑回归. 计算机研究与发展, 2019, **56**(10): 2243−2249)

24 Phong L T, Aono Y, Hayashi T, Wang L H, Moriai S. Privacy-preserving deep learning via additively homomorphic encryption. *IEEE Transactions on Information Forensics and Security*, 2018, **13**(5): 1333−1345

25 Ou W, Zeng J H, Guo Z J, Yan W Q, Liu D W, Fuentes S. A homomorphic-encryption-based vertical federated learning scheme for rick management. *Computer Science and Information Systems*, 2020, **17**(3): 819−834

26 Chen H, Chillotti I, Song Y. Improved bootstrapping for approximate homomorphic encryption. In: Proceedings of the 38th Annual International Conference on the Theory and Applications of Cryptographic Techniques. Darmstadt, Germany: Springer, 2019. 34−54

27 Xiao X D, Wu T, Chen Y F, Fan X Y. Privacy-preserved approximate classification based on homomorphic encryption. *Mathematical and Computational Applications*, 2019, **24**(4): Article No. 92

28 Zhang Z H, Yao F, Gao T G. A hybrid image encryption algorithm based on chaos system and simplified advanced encryption system. *International Journal of Multimedia Data Engineering and Management (IJMDEM)*, 2020, **11**(4): Article No. 1

29 Luo Y Q, Yu J, Lai W R, Liu L F. A novel chaotic image encryption algorithm based on improved baker map and logistic map. *Multimedia Tools and Applications*, 2019, **78**(15): 22023−22043

30 Sathiyamurthi P, Ramakrishnan S. Speech encryption algorithm using FFT and 3D-Lorenz-logistic chaotic map. *Multimedia Tools and Applications*, 2020, **79**(25): 17817−17835

31 Sattler F, Müller K, Samek W. Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints. *IEEE Transactions on Neural Networks and*

Learning Systems, 2020, **32**(8): 3710−3722

32 Al-Sharman M, Murdoch D, Cao D P, Lv C, Zweiri Y, Rayside D, et al. A sensorless state estimation for a safety-oriented cyber-physical system in urban driving: Deep learning approach. *IEEE/CAA Journal of Automatica Sinica*, 2021, **8**(1): 169−178

33 Weng J S, Weng J, Zhang J L, Li M, Zhang Y, Luo W Q, et al. DeepChain: Auditable and privacy-preserving deep learning with blockchain-based incentive. *IEEE Transactions on Dependable and Secure Computing*, 2021, **18**(5): 2438−2455

34 Sattler F, Wiedemann S, Müller K R, Samek W. Robust and communication-efficient federated learning from non-i.i.d. data. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, **31**(9): 3400−3413

35 Xu G W, Li H W, Zhang Y, Xu S M, Ning J T, Deng R H. Privacy-preserving federated deep learning with irregular users. *IEEE Transactions on Dependable and Secure Computing*, 2022, **19**(2): 1364−1381

36 Teng S H, Wu N Q, Zhu H B, Zhang W. SVM-DT-based adaptive and collaborative intrusion detection. *IEEE/CAA Journal of Automatica Sinica*, 2018, **5**(1): 108−118

张泽辉 南开大学软件学院博士研究生. 2019 年获得武汉理工大学硕士学位. 主要研究方向为联邦学习, 故障诊断和智能船舶控制.
E-mail: zhangtianxia918@163.com
(**ZHANG Ze-Hui** Ph.D. candidate at the College of Software, Nankai University. He received his master degree from Wuhan University of Technology in 2019. His research interest covers federated learning, fault diagnosis and intelligent ship control.)

李庆丹 南开大学软件学院硕士研究生. 主要研究方向为图像加密, 信息安全.
E-mail: lqd18812745024@163.com
(**LI Qing-Dan** Master student at the College of Software, Nankai University. Her research interest covers image encryption and information security.)

富 瑶 南开大学软件学院硕士研究生. 主要研究方向为云端数据完整性验证, 信息安全.
E-mail: fuyao_tj@163.com
(**FU Yao** Master student at the College of Software, Nankai University. Her research interest covers cloud data integrity verification and information security.)

何宁昕 南开大学软件学院硕士研究生. 2020 年获得河北经贸大学学士学位. 主要研究方向为信息安全, 联邦学习.
E-mail: ningxinhe1998@163.com
(**HE Ning-Xin** Master student at the College of Software, Nankai University. She received her bachelor degree from Hebei University of Economics and Business in 2020. Her research interest covers information security and federated learning.)

高铁杠 南开大学软件学院教授. 1991 年获华中理工大学应用数学专业硕士学位, 2005 年获南开大学博士学位. 主要研究方向为联邦学习, 图像水印, 信息隐藏和云端数据安全. 本文通信作者.
E-mail: gaotiegang@nankai.edu.cn
(**GAO Tie-Gang** Professor at the College of Software, Nankai University. He received his master degree in applied mathematics from Huazhong University of Science and Technology in 1991, and Ph.D. degree from Nankai University in 2005. His research interest covers federated learning, image watermarking, information hiding and cloud data security. Corresponding author of this paper.)