ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 39,No. 6,Nov. 2024,pp. 1532－1542
DOI:10. 16337/j. 1004-9037. 2024. 06. 020

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

基于智能合约和联邦存储的异步联邦学习模型

刘星辰[1,2]，杜军平[1,2]，梁美玉[1,2]，李　昂[1,2]

(1. 北京邮电大学计算机学院(国家示范性软件学院),北京 100876;2. 智能通信软件与多媒体北京市重点实验室(北京邮电大学),北京 100876)

摘　要：公共安全突发事件中对数据安全的重视程度越来越高,联邦学习由于不再需要上传数据到中心服务器进行计算,减少了隐私泄露的可能而受到广泛关注。然而当前基于智能合约的联邦学习由于运算较大,存在着效率低等缺陷,因此本文提出了一种面向公共卫生突发事件检测的智能合约与联邦存储的异步联邦学习方法。该方法允许联邦节点在任何时间加入和退出联邦学习;依托智能合约与分布存储,进一步增加了公共卫生安全领域的数据安全与训练效率;同时采用自适应的差分隐私对其上传到分布式存储节点的梯度进行动态保护,进一步降低了隐私泄露的风险。在公共数据集和公共卫生安全数据集上大量的实验表明,本文提出的方法在精度上优于已知的对比方法,且在达到相同精度的情况下所需时间更少。

关键词：智能合约;联邦学习;公共卫生突发事件;联邦存储

中图分类号：TP391　　**文献标志码**：A

Asynchronous Federated Model of Public Health Emergency Monitoring Based on Smart Contract and Federated Storage

LIU Xingchen[1,2], DU Junping[1,2], LIANG Meiyu[1,2], LI Ang[1,2]

(1. School of Computer Science (National Pilot School of Software Engineering), Beijing University of Posts and Telecommunications, Beijing 100876, China; 2. Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia (Beijing University of Posts and Telecommunications), Beijing 100876, China)

Abstract: With the increasing emphasis on data security in public safety emergencies, federated learning has gained attention for its ability to perform computations without uploading data to a central server, thereby reducing the risk of privacy breaches. However, current federated learning approaches based on smart contracts face challenges such as inefficiency due to their computational demands. To address it, this paper proposes an asynchronous federated learning method for detecting public health emergencies, integrating smart contracts and federated storage. This approach allows federated nodes to join and leave the federated learning process at any time. By leveraging smart contracts and distributed storage, it enhances data security and training efficiency in the public health domain. Furthermore, adaptive differential privacy is employed to dynamically protect the gradients uploaded to distributed storage nodes, further reducing the risk of privacy leakage. Extensive experiments conducted on public datasets and public health security datasets demonstrate that the proposed method outperforms existing approaches in terms of

基金项目:国家自然科学基金 (U22B2038,62192784,62172056)。
收稿日期:2023-06-29;**修订日期**:2023-10-10

accuracy and requires less time to achieve the same level of precision.

Key words: smart contract; federated learning; public health emergency; federated storage

引 言

机器学习已经成为了当代人工智能技术的核心。然而,传统的机器学习方法需要大量的数据来训练可用的模型,这往往会面临数据保护法的限制以及数据泄露的风险。近年来,随着计算设备和互联网的普及,分布式设备中所存储的数据量越来越大、更具多样性[1-3],而这需要更加复杂的数据收集和资源分配。为了应对新环境的挑战,联邦学习被引入到了机器学习领域[4-8]。联邦学习能够使得不同参与者之间协作训练一个机器学习模型,而不需要在此过程中明文传输和共享自己本地的数据。每个参与者都只需要访问本地数据,并且在本地完成第一轮训练后,上传与各自相对应的更新参数信息来实现全局的优化。联邦学习的这种方式有助于提高分布式机器学习的效率,并降低了数据泄露的风险[9-10]。联邦学习的研究发展已为社会和工业界提供了广泛的应用前景,涉及的领域包括图像分类、语音识别以及自然语言处理等。

联邦学习是一种通过中央服务器聚合每个设备上传的模型或参数进行模型更新。因此,联邦学习不需要设备共享数据,既能够完成全局模型的更新,又不会泄露数据隐私[11]。目前联邦学习主要研究热点是聚合方法使模型收敛性更好以及隐私保护方法的研究等。联邦平均是常用的联邦学习方法之一[12],即通过各个分布式设备来更新全局模型,并将更新的模型或相关参数上传到中央服务器上进行平均。这种方法的主要挑战是单点故障和不确定性[13]。由于中央服务器是单一的,如果出现故障,整个训练过程将失败。此外,各个参与方之间可能互不信任,这也增加了使用该方法的不确定性。虽然该方法在保护数据隐私方面取得了很大进展,但仍然需要进一步商榷和探索,以构建更加鲁棒和安全的联邦学习系统。研究人员正在积极寻找新的解决方案,以应对人工智能模型训练中的种种挑战。他们考虑采用去中心化的方法[14],将任务分配给多个节点进行处理,以消除单点故障的风险。此外,他们还在探索使用安全多方计算等技术,以确保参与者之间的数据共享安全可靠,从而提高模型训练的可靠性和安全性。

智能合约[15]通过预先编程的交易规则来自动触发交易。不过,由于智能合约需要在执行环境中进行操作,因此并没有得到广泛应用。随着以太坊、物联网应用(Internet of things application,IOTA)等区块链技术的出现,为智能合约的实现和应用提供了更多的可能性[16-23]。智能合约是一种通过编写明确的条件语句来实现自动执行的技术,与传统合约相比具有许多优势。首先,智能合约公开透明[24],避免了因理解不同而产生的分歧和纠纷[25]。其次,交易过程中合约无法被篡改或被拒绝,从而降低了人力资源成本并提高了交易效率。最后,存储在分布式区块链账户中的智能合约不会因为电力故障、节点故障等原因而出错或遗失[26]。智能合约作为一种成熟的去中心化技术,在联邦学习等领域也得到了广泛应用。由于智能合约的优势,很多工作使用区块链技术实现去中心化的联邦学习。这种方法可以克服传统联邦学习中的单点故障风险[27],即使某个客户端在因本身发生连接超时的情况下,整个联邦体系仍然可以正常地聚合其余客户端传递来的梯度信息。但是由于客户端的不可控性和区块链带来的机制限制,这种方法可能会受到运算复杂度较高和不同版本间梯度差异难以收敛的限制[28]。同时异步的更新会使网络性能好、频繁聚合的客户端拥有更大的权重,全局模型更偏向传输快的客户端,而不是全局的最优解。

本文提出了一种基于智能合约与分布式存储框架的联邦学习方法(Smart contact and federated storage based asynchronous federated learning,SCFS-AFL)。为了减少因为等待个别慢客户端而造成的

等待时间,SCFS-AFL采用了异步聚合策略,同时为了减少长时间未聚合的版本对当前全局模型产生负向影响,对其进行了权重衰减的策略。采用自适应差分隐私保护的方式,对上传到分布式文件存储系统的梯度进行保护。此外,为了使SCFS-AFL同样适用于大型模型,减少梯度传递的时间,本文提出了分布式存储策略。这个策略将模型数据与对应的哈希值分隔开,并将所有的智能合约用于联邦学习,保障分布式节点的安全性。本文在公共卫生安全数据集和公开数据集上分别训练多个模型来评估SCFS-AFL。实验结果表明,本文提出的聚合方法可以在解决传统联邦学习中单点故障的同时达到相似的模型性能,具有更强的鲁棒性。

1 相关工作

联邦平均算法[13]是联邦学习领域的开篇之作。在这个算法中,一个中央服务器负责维护全局模型,每轮开始时将全局模型发送给选定的参与方,然后将这些参与方上传的模型或更新进行聚合。虽然联邦平均算法已经被应用到金融、医疗和推荐等领域,但是它的应用受到非独立同分布数据的限制。为了解决这个问题,文献[29]提出在联邦平均算法中增加一个近端项,以限制更新的差异,从而缓解非独立同分布数据对联邦学习的影响。此外,为了克服联邦学习中的异构挑战,进一步提高联邦学习的可靠性和鲁棒性,需要探索新的联邦学习架构,这些架构不仅可以处理非独立同分布数据,还可以保证训练的连续性和可扩展性[30]。

在联邦学习中,单点故障一直是一个重要的问题。为了解决这个问题,目前有一些使用区块链技术实现去中心化联邦学习的方法[31-33],这样即使出现节点故障,系统的运行也不会受到影响,因为历史模型仍然存在于区块链账本中[34]。另外一种方法是使用带智能合约的私有以太坊网络来存储模型的副本[35],并聚合其他参与者发送的模型更新。这种方法的优势在于可以利用智能合约来进行更加复杂的模型更新操作。然而,引入智能合约也带来了隐私泄露的风险,因此需要使用不同的加密技术来保护数据隐私。一些通用的加密技术包括对称加密、非对称加密、同态加密和差分隐私。这些技术可以用来加密联邦学习中的数据和模型,从而保护数据隐私并提高系统的安全性。在模型参数比较大的情况下,上述策略会暴露出明显的运行效率较低的弊端,无法满足日益增长的数据量以及复杂的模型结构[36-38]。文献[39]提出可以用基于哈希量化的点对点分布式文件系统,与智能合约结合使用。

联邦学习作为一种分布式机器学习方法被广泛应用于数据隐私保护领域。联邦学习的核心在于,多个参与方协同训练一个共同的机器学习模型,避免将数据集中到一个中心化的位置,从而减少了数据隐私泄露的风险。然而,在联邦学习中,由于不同参与方的数据分布和特征不同,模型训练可能会受到一些限制和挑战。为此,隐私保护技术成为联邦学习中的一个热点问题[40]。

目前,针对联邦学习的隐私保护,差分隐私、同态加密和安全多方计算等技术被广泛应用[41-44]。差分隐私技术通过给客户端添加噪声的方式来保护该客户端的隐私,但会对模型精度产生影响。叶青青等[45]使用差分隐私来保护模型传输的隐私,但可能会影响模型的准确性。一些研究者通过梯度裁剪来减缓隐私的泄露:Bun等[46]采用线性上限$a(\lambda)$对梯度进行裁剪,以限制联邦客户端数据的泄露;Truex等[47]使用客户端本地扰动数据,并使用Paillier加密算法来聚合扰动后的数据,以有效保护隐私,这种方法可能导致训练时间较长和通信开销较大;Hardy等[48-52]结合实体解析和同态加密来处理纵向分布模型数据。

联邦学习中的隐私保护应根据不同的场景和要求选择合适的隐私保护种类[53-56]。总的来说,联邦学习隐私保护是一个研究热点,随着技术的不断发展,未来可能会出现更多有效的隐私保护技术,以更好地应对不同场景和需求的挑战。

2 算法设计

本文提出的面向公共卫生突发事件检测的SCFS-AFL方法框架如图1所示。其采用了智能合约的思想,在异步联邦学习过程中实现全局联邦模型,并增加了抗攻击能力和安全性。参与者需至少在1个节点托管智能合约地址,执行模型训练和上传等过程。智能合约在所有部署的节点上运行,保证在节点发生故障时仍可持续工作。该方法采用异步聚合方式,允许任意时刻加入并上传模型更新,并不要求最少的参与者数量。同时,该方法引入联邦隐私保护,对模型进行自适应的差分隐私保护,从而保护数据隐私。

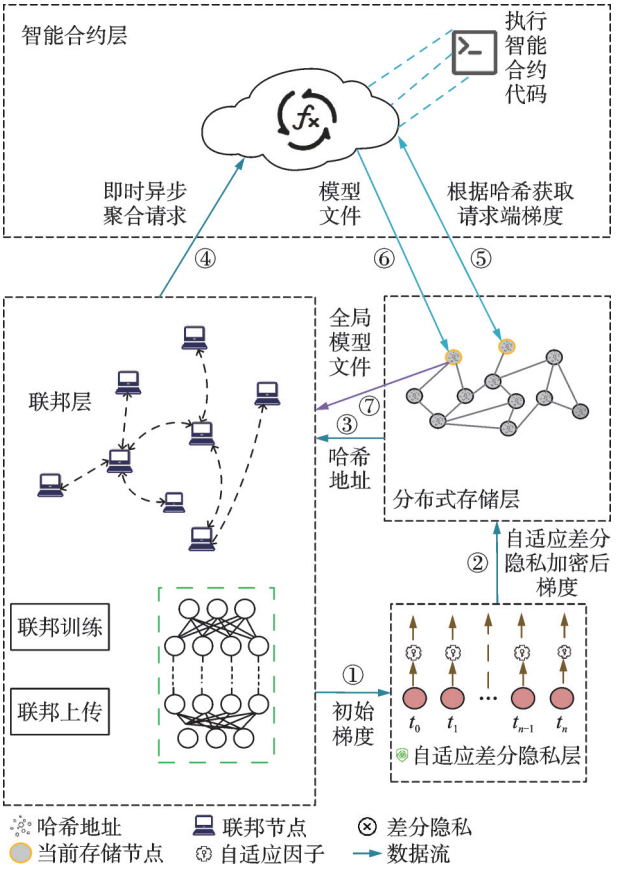

图1 SCFS-AFL示意图

Fig.1 Illustration of SCFS-AFL

SCFS-AFL方法的实现过程如下:初始化全局模型,联邦各方在本地执行梯度下降算法更新模型,上传到分布式文件存储系统中。在本地联邦节点传输到分布式文件系统的过程中,需要对其进行自适应差分隐私的联邦隐私保护。当参与者的模型聚合到中心模型后,对其进行自适应的异步联邦聚合策略,聚合这些更新并得到一个新的全局模型。在这个过程中,智能合约节点依据模型哈希信息进行存储和访问。

SCFS-AFL方法的优点是在保护数据隐私的同时提高安全性和抗攻击能力。此外,由于该方法不要求最少参与者数量,因此可以增强联邦学习的灵活性和鲁棒性。

2.1 基于智能合约自适应的的异步联邦聚合策略

本文介绍了使用异步方法进行联邦学习的一种模型更新方法。与典型的同步算法不同,该模型更新方法如式(1)所示。

$$\theta_{t+1}=\frac{a(n_t-d_t)\theta_t+(1-a)(n_t^{\text{new}}+d_t)\theta_t^*}{n_{t+1}} \tag{1}$$

式中:θ_t表示第t轮模型,n_t表示用第t轮用于模型训练的样本总数,n_t^{new}表示新样本数量,a(取值0~1)为调节因子,d_t为第$t+1$轮使用前t轮数据量;θ_t^*表示第t轮梯度,这里表示第$t+1$轮的参数是由第t轮的参数和第t轮的梯度通过数据量等参数得到。当接收到客户传递来的梯度时,运行生成的工作量证明(Proof of work, POW)算法,生成块进行存储。

2.2 基于哈希量化的质效均衡联邦通信优化

哈希量化是一种可以将梯度上传和下载速度进行优化的方法。在异步联邦学习中,智能合约作为类似中心节点,分布式节点将梯度文件上传到智能合约上。然后智能合约控制刚上传的梯度立即与全局模型进行聚合,然后下发给当前的联邦节点。这个过程一般需要比较长的时间,因为上传和下载梯

度文件需要消耗大量的时间,尤其是在大规模的异步联邦学习中。

哈希量化被引入到异步联邦学习中。该方法首先使用一个分布式的文件系统来存储梯度文件。在上传梯度文件之前,梯度文件会先被哈希计算,并且分布式文件系统会给出一个哈希值,然后该节点将梯度文件和这个哈希值一起以文件的形式上传到分布式文件系统上。当有节点需要下载梯度时,会向智能合约节点请求下载梯度。智能合约节点会根据节点传来的哈希值从分布式文件系统中找到相应的文件进行下载,然后与全局模型聚合。这个方式是以哈希计算为基础的,因此被称为哈希量化。哈希量化可以大大减少上传和下载梯度文件所需的时间和网络带宽。这是因为哈希计算是一个比较快速和高效的操作,其结果可以用来验证梯度的完整性和准确度。每个节点只需上传和下载少量的数据,因此可以大大减少节点之间的通信开销。哈希量化可以优化异步联邦学习中梯度上传和下载的速度及网络带宽,加快联邦学习的速度,并为跨设备、跨机构的合作提供可能性。

2.3 基于自适应差分隐私的联邦隐私保护算法

本节介绍了一种基于AdaMod算法设计的自适应差分隐私优化算法,具体描述如表1所示。其采用随时间降低噪声强度的方法自适应添加噪声,可以保护神经网络模型训练数据集的隐私,并降低隐私损失。相对于其他非自适应更新学习速率的差分隐私优化算法,该算法的训练迭代次数更少,隐私

表1 自适应的差分隐私优化算法

Table 1 Adaptive differential privacy optimization algorithm

自适应差分隐私优化算法
输入:样本$\{x_1,x_2,\cdots,x_N\}$,学习率$\{\eta_t\}_{t=1}^{B}$,$t=1$表示当前第1轮,总共B轮,噪声参数σ(即随机的高斯噪声,$\sigma_t=(\sigma_t-1)*0.99$,其中0.99是衰减常数,下衰减同理),$\sigma$随着迭代的增加会逐渐减少,线性衰减$\{\beta_1,\beta_2,\beta_3\}$,梯度修剪范数$C$,迭代次数$T$,损失函数$L(\theta)=\frac{1}{N}\sum_i L(\theta,x_i)$。
初始化:随机初始化参数θ_0,初始化参数$m_0=0,v_0=0,s_0=0$(m表示一阶动量,v表示二阶动量,s表示中间变量)
for $t\subseteq[T]$ do
以概率L/N随机采样一组样本S_t
步骤1:梯度计算:对于每个样本$i\in S_t$,计算梯度$g_t(x_i)\leftarrow\nabla_{\theta_t}L(\theta_t,x_i)$
步骤2:梯度裁剪:$g_t(x_i)\leftarrow\frac{g_t(x_i)}{\max\left(1,\|g_t(x_i)\|_2/C\right)}$
步骤3:计算学习速率
计算累计梯度:$g\leftarrow\sum_{i=0}^{m}g_t(x_i)$
$m_t=\beta_1 m_{t-1}+(1-\beta_1)g_t$
$v_t=\beta_2 v_{t-1}+(1-\beta_2)g_t^2$
$m_t=m_t/(1-\beta_1^t)$
$v_t=v_t/(1-\beta_2^t)$
$s_t=\beta_3 s_{t-1}+(1-\beta_3)\eta_t$
$\eta_t=\min(\eta_t,s_t)$
步骤4:噪声添加$g_t\leftarrow\frac{1}{L}\sum_i m_t+N(0,\sigma_t^2C^2I)$
步骤5:参数更新$\theta_t=\theta_{t-1}-\eta_t g_t$
输出:θ_T

损失更小。此外,该算法继承了AdaMod算法的优越性,使得模型训练更加灵活,缓解了差分隐私优化算法调参困难的问题。

计算学习速率的过程基于梯度累积值的一阶矩和二阶矩估计值,并计算当前时刻的指数滑动平均值,最终选取较小值作为当前的学习速率。这个过程可以帮助控制模型参数的更新速率。

在差分隐私场景下,为了防止攻击者通过模型输出判断训练数据集中的某一个数据,会在最终累积的梯度值中添加高斯噪声。噪声的大小由梯度修剪阈值和噪声参数决定,其中梯度修剪阈值限制梯度向量的值,而噪声参数控制噪声的大小。

3 实验设置

本文通过以下3点研究问题来引导实验的设置:(1)面向公共卫生突发事件检测的智能合约与联邦存储的异步联邦学习方法表现能否优于前沿的联邦学习算法;(2)SCFS-AFL方法的主要组成部分对于联邦学习任务是否存在贡献;(3)SCFS-AFL方法是否在运行效率上优于传统联邦学习的模型。

本文将SCFS-AFL与相关的基准方法和前沿方法进行比较,对比方法分别是集中式训练、FedAvg[13]、BlockFL[57]、TrustFed[58]以及DAA-FL[59]。

在经典的MNIST数据集与公共卫生安全数据集对比不同方法的训练精度和训练时的通信开销,两种数据集规格如表2所示。在门户网站爬取的公共安全文本数据集进行事件检测。在Linux平台上进行实验,并利用Python3脚本实现。本文用的DNN模型是1个3层的网络模型,激活函数为Relu,为了防止过拟合,采用10%的Dropout。本文设计一共有100个客户端,客户端对图像进行训练时采用Adam优化器。

表2 实验的两种数据集

Table 2 Two data sets for experiment

数据集	训练/测试样本数	标签数	描述
公共卫生安全数据集	3 000/1 000	2	公共卫生突发事件文本分类
MNIST数据集	50 000/10 000	10	手写数字识别

4 实验结果

4.1 模型精度

本文选取分类准确度F_1、分类精度Percision和召回率Recall作为评价指标。图2显示了在公共安全数据集上的Loss曲线,从图中下降的趋势可以看出,本文提出的SCFS-AFL方法可以有效地进行联邦学习。表3、4以自然语言处理模型为主干模型,在公共安全数据集上的效果相比于BlockFL也领先1.1%。实验结果表明,SCFS-AFL都可以得到接近集中训练方法的模型精度。表5展示了在MNIST数据集上的实验结果。从结果看出,集中式学习由于只有1个客户端,没有独立同分布的限制,能够取得更好的结果,相比于FedAvg有着更好的效果。在联邦场景下,相比于其他方法,本文提出的异步方法由于

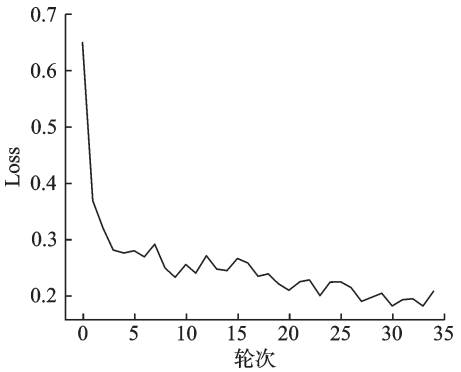

图2 公共安全数据集Loss下降图

Fig.2 Loss decline for public safety dataset

表3 对比方法在公共卫生安全数据6层BERT指标对比

Table 3 Comparison of BERT indicators in six layers of public health safety data using comparative methods

对比方法	5轮 F_1	5轮 Percision	5轮 Recall	10轮 F_1	10轮 Percison	10轮 Recall
集中式	0.869	0.905	0.835	0.891	0.925	0.859
FedAvg	0.823	0.855	0.793	0.835	0.876	0.797
BlockFL	0.829	0.858	0.801	0.835	0.875	0.798
TrustFed	0.830	0.860	0.800	0.838	0.880	0.800
DAA-FL	0.832	0.865	0.802	0.842	0.883	0.804
SCFS-AFL	**0.835**	**0.868**	**0.804**	**0.844**	**0.885**	**0.806**

表4 对比方法在公共卫生安全数据12层BERT指标对比

Table 4 Comparison of BERT indicators in 12 layers of public health safety data using comparative methods

对比方法	5轮 F_1	5轮 Percision	5轮 Recall	10轮 F_1	10轮 Percison	10轮 Recall
集中式	0.870	0.926	0.819	0.932	0.934	0.930
FedAvg	0.854	0.870	0.838	0.883	0.882	0.883
BlockFL	0.866	0.882	0.852	0.878	0.898	0.859
TrustFed	0.868	**0.884**	0.853	0.880	0.900	0.859
DAA-FL	0.872	0.883	0.862	0.871	0.903	0.841
SCFS-AFL	**0.873**	0.882	**0.864**	**0.874**	**0.909**	**0.843**

表5 SCFS-AFL在MNIST数据集上的性能

Table 5 SCFS-AFL performance on MNIST dataset

对比方法	5轮 F_1	5轮 Percision	5轮 Recall	10轮 F_1	10轮 Percison	10轮 Recall
集中式	0.943	0.954	0.932	0.970	0.975	0.965
FedAvg	0.726	0.756	0.698	0.786	0.812	0.762
BlockFL	0.755	0.793	0.720	0.811	0.834	0.789
TrustFed	0.758	0.798	0.723	0.813	0.835	0.790
DAA-FL	0.782	0.811	0.755	0.851	0.862	0.840
SCFS-AFL	**0.786**	**0.812**	**0.762**	**0.853**	**0.865**	**0.841**

有更多的聚合次数,而有着更好的指标。在3层DNN的网络结构和MNIST数据集的情况下,相比于FedAvg在10轮 F_1 有6.7%的提升。

4.2 通信效率分析(消融实验)

图3分别展示了3种对比方法及本文提出的异步联邦算法以及异步联邦算法和哈希优化的效率分析图。在MNIST数据集上进行时间分析,本文提出的方法运行效率也优于其他联邦学习方法。大量的实验结果证明了SCFS-AFL的有效性和高效性。

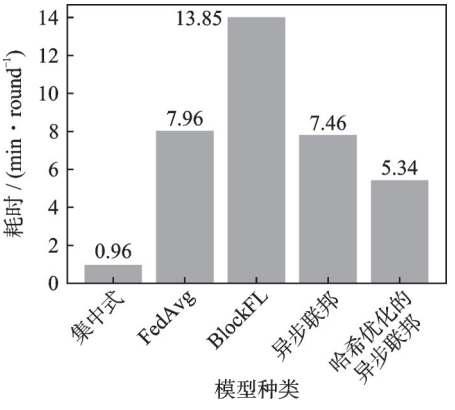

图3 联邦学习算法聚合效率图

Fig.3 Aggregation efficiency graph for federated learning algorithm

4.3 自适应差分隐私对比实验

DPAdam是一种差分隐私自适应学习速率的算法,原理是通过减少聚合次数,从而减少差分隐私对联邦学习性

能的影响。图4和图5分别展示了在高(噪声系数$\varepsilon=1$)、低(噪声系数$\varepsilon=0.5$)两种隐私等级下本文提出的自适应差分隐私算法和DPAdam算法准确度Acc的对比图。实验采用公共安全数据集,梯度修剪阈值为$C=0.1$、噪声参数$\sigma=0.05$、初始学习速率$\eta=0.000\,1$。

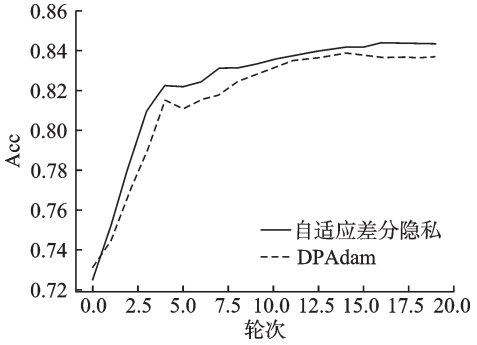

图4 高隐私保护情况下Acc对比图

Fig.4 Acc comparison chart under high privacy protection

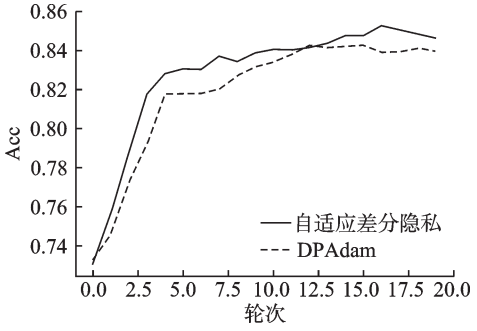

图5 低隐私保护情况下Acc对比图

Fig.5 Acc comparison chart under low privacy protection

从图4,5可以看出,在高、低两个隐私等级下,本文提出的自适应差分隐私算法优于DPAdam算法。此实验也说明了在BERT等大模型联邦场景下需要与差分隐私结合的场景下,相对于学习速率固定的梯度更新算法,学习速率随更新情况自适应变化的梯度更新算法能够更好地权衡模型的隐私性和准确度。同时本文提出的自适应差分隐私算法在一定程度上可以有效防止出现学习速率过大的情况,保证深度学习模型在加入拉普拉斯噪声或者高斯噪声的情况下仍可以保持较好的性能。

4.4 算法复杂度分析

联邦学习耗时主要在梯度传输的通信开销,其余的时间消耗在模型的训练以及模型参数的噪声上。通过上文对通信效率的分析,由于采用了分布式存储系统,本文提出的异步联邦算法在通信效率优于FedAvg为代表的同步式联邦学习算法以及需要工作量证明的基于区块链的联邦学习算法。其次本文采取的联邦学习隐私保护算法为差分隐私的隐私保护算法,相比于一些基于对称加密和非对称加密的算法,在添加噪声处可以通过添加DNN层的方式来实现,几乎不会增加时间复杂度。在模型训练过程中,由于采用基础的模型结构,所以在模型训练过程中和普通的联邦学习的时间复杂度相似,同时时间复杂度也与数据集无关。在训练过程中,由于模型的文件以及梯度需要加载到内存中,其空间复杂度和单机模型训练时类似。对整个联邦系统来说,本文提出的异步聚合方法通过联邦存储将模型文件映射到训练节点所在的文件上,可以减少整个联邦系统的存储的压力,亦可得出所采用的联邦深度学习算法空间复杂度与数据集无关的结论。

5 结束语

本文提出了一种基于智能合约的异步联邦学习方法,采用哈希量化的方式优化智能合约过程中效率不高的问题,同时提出一种自适应差分隐私保护算法。在MNIST数据集以及公共卫生安全数据集上相较于前沿算法取得了较好的指标,同时在相同的保护力度上,指标效果也高于目前常见的隐私保护算法。未来的工作将聚焦于联邦学习上的质效均衡,优化联邦学习在真实场景下的效率与模型效果。

参考文献：

[1] 谭作文,张连福.机器学习隐私保护研究综述[J].软件学报,2020, 31(7): 2127-2156.
TAN Zuowen, ZHANG Lianfu. A review of machine learning privacy protection research[J]. Journal of Software, 2020, 31(7): 2127-2156

[2] KOU Feifei, DU Junping, YANG Congxian, et al. Hashtag recommendation based on multi-features of microblogs[J]. Journal of Computer Science and Technology, 2018, 33(4): 711-726.

[3] MENG Deyuan, JIA Yingmin, DU Junping, et al. Tracking algorithms for multiagent systems[J]. IEEE Transactions on Neural Networks and Learning Systems, 2013, 24(10): 1660-1676.

[4] LI Yawen, LI Wenling, XUE Zhe. Federated learning with stochastic quantization[J]. International Journal of Intelligent Systems, 2022, 37(12): 11600-11621.

[5] GUAN Z, LI Y, XUE Z, et al. Federated graph neural network for cross-graph node classification[C]//Proceedings of 2021 IEEE 7th International Conference on Cloud Computing and Intelligent Systems (CCIS). [S.l.]: IEEE, 2021: 418-422.

[6] LI Ang, DU Junping, KOU Feifei, et al. Scientific and technological information oriented semantics-adversarial and media-adversarial cross-media retrieval[EB/OL]. (2022-03-16)[2023-08-01]. https://arxiv.org/abs/2203.08615.

[7] CAO Tengfei, XU Changqiao, DU Junping, et al. Reliable and efficient multimedia service optimization for edge computing-based 5G networks: Game theoretic approaches[J]. IEEE Transactions on Network and Service Management, 2020, 17(3): 1610-1625.

[8] CAO J, MAO D H, CAI Q, et al. A review of object representation based on local features[J]. Journal of Zhejiang University Science C, 2013, 14(7): 495-504.

[9] ZHANG Chen, XIE Yu, BAI Huang, et al. A survey on federated learning[J]. Knowledge-Based Systems, 2021, 216: 106775.

[10] MA Xiaodong, ZHU Jia, LIN Zhihao, et al. A state-of-the-art survey on solving non-IID data in federated learning[J]. Future Generation Computer Systems, 2022, 135: 244-258.

[11] BOOPALAN P, RAMU S P, PHAM Q V, et al. Fusion of federated learning and industrial internet of things: A survey[J]. Computer Networks, 2022, 212(1): 109048.

[12] CAI Hongyun, ZHANG Yu, WANG Shiyun, et al. Trusted federation security aggregation algorithm based on similarity clustering[J]. Journal of Electronics and Information, 2023, 45: 11.

[13] MCMAHAN B, MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data [C]//Proceedings of Artificial Intelligence and Statistics. [S.l.]: PMLR, 2017: 1273-1282.

[14] WANG R, TSAI W T T. Asynchronous federated learning system based on permissioned blockchains[J]. Sensors, 2022, 22(4): 1672.

[15] 胡甜媛,李泽成,李必信,等.智能合约的合约安全和隐私安全研究综述[J].计算机学报, 2021, 44(12): 2485-2514.
HU Tianyuan, LI Zecheng, LI Bixin, et al. A review of contract security and privacy security research on smart contracts[J]. Chinese Journal of Computers, 2021, 44(12): 2485-2514

[16] KIRLI D, COURAUD B, ROBU V, et al. Smart contracts in energy systems: A systematic review of fundamental approaches and implementations[J]. Renewable and Sustainable Energy Reviews, 2022, 158: 112013.

[17] BALCERZAK A P, NICA E, ROGALSKA E, et al. Blockchain technology and smart contracts in decentralized governance systems[J]. Administrative Sciences, 2022, 12(3): 96.

[18] LIN S Y, ZHANG L, LI J, et al. A survey of application research based on blockchain smart contract[J]. Wireless Networks, 2022, 28(2): 635-690.

[19] TOMASZ G. Reconfigurable smart contracts for renewable energy exchange with reuse of verification rules[J]. Applied Sciences, 2022, 12(11): 5339.

[20] KUSHWAHA S S, JOSHI S, SINGH D, et al. Systematic review of security vulnerabilities in ethereum blockchain smart contract[J]. IEEE Access, 2022, 10: 6605-6621.

[21] ZHOU H, MILANIFARD A, MAKANJU A. The state of Ethereum smart contracts security: Vulnerabilities, countermeasures, and tool support[J]. Journal of Cybersecurity and Privacy, 2022, 2(2): 358-378.

[22] MENG Deyuan, JIA Yingmin, DU Junping, et al. High-precision formation control of nonlinear multi-agent systems with

switching topologies: A learning approach[J]. International Journal of Robust and Nonlinear Control, 2015, 25(13): 1993-2018.

[23] LI Yawen, JIA Yingmin, DU Junping. Resilient filtering for nonlinear complex networks with multiplicative noise[J]. IEEE Transactions on Automatic Control, 2018, 64(6): 2522-2528.

[24] LENNART A. The non-fungible token (NFT) market and its relationship with Bitcoin and Ethereum[J]. FinTech, 2022, 1(3): 216-224.

[25] LIU L, TSAI W T, BHUIYAN M Z A, et al. Blockchain-enabled fraud discovery through abnormal smart contract detection on Ethereum[J]. Future Generation Computer Systems, 2022, 128: 158-166.

[26] ANTEVSKI K, CARLOS J. Federation in dynamic environments: Can blockchain be the solution[J], IEEE Communications Magazine, 2022, 60(2): 32-38.

[27] OUYANG L W, WANG X, TIAN Y L, et al. Artificial identification: A novel privacy framework for federated learning based on blockchain[J]. IEEE Transactions on Computational Social Systems, 2023, 10(6): 3576-3585.

[28] YANG Qiang. Overview of federated learning algorithms in group intelligence[J]. Journal of Intelligent Science and Technology, 2022, 4(1): 29-44.

[29] 顾育豪,白跃彬.联邦学习模型安全与隐私研究进展[J].软件学报, 2023, 34(6): 2833-2864.
GU Yuhao, BAI Yuebin. Research progress on security and privacy of federated learning models[J]. Journal of Software, 2023, 34(6): 2833-2864.

[30] LI TIAN, SAHU A K, ZAHEER M, et al. Federated optimization in heterogeneous networks[J]. Proceedings of Machine Learning and Systems, 2020, 2: 429-450.

[31] KNOWLES C L, D'AGOSTINO S R, KUNZE M G, et al. A systematic review of asynchronous online learning opportunities for paraeducators[J]. The Journal of Special Education, 2022, 56(3): 168-178.

[32] WEI Xinlei, DU Junping, LIANG Meiyu, et al. Boosting deep attribute learning via support vector regression for fast moving crowd counting[J]. Pattern Recognition Letters, 2019, 119: 12-23.

[33] LI Wenling, JIA Yingmin, DU Junping. Tobit Kalman filter with time-correlated multiplicative measurement noise[J]. IET Control Theory & Applications, 2017, 11(1): 122-128.

[34] RAMANAN P, NAKAYAMA K. Baffle: Blockchain based aggregator free federated learning[C]//Proceedings of 2020 IEEE International Conference on Blockchain (Blockchain). [S.l.]: IEEE, 2020: 72-81.

[35] KIM H, PARK J, BENNIS M, et al. Blockchained on-device federated learning[J]. IEEE Communications Letters, 2019, 24(6): 1279-1283.

[36] XIAO Shitao, SHAO Yingxia, LI Yawen. LECF: Recommendation via learnable edge collaborative filtering[J]. Science China Information Sciences, 2022, 65(1): 1-15.

[37] SHAO Yingxia, HUANG Shiyue, LI Yawen, et al. Memory-aware framework for fast and scalable second-order random walk over billion-edge natural graphs[J]. The VLDB Journal, 2021, 30(5): 769-797.

[38] LI Yawen, YUAN Ye, WANG Yishu, et al. Distributed multimodal path queries[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(7): 3196-3210.

[39] SUBRAMANIAN H, SUBRAMANIAN S. Improving diagnosis through digital pathology: Proof-of-concept implementation using smart contracts and decentralized file storage[J]. Journal of Medical Internet Research, 2022, 24(3): e34207.

[40] KASYAP H, MANNA A, TRIPATHY S. An efficient blockchain assisted reputation aware decentralized federated learning framework[J]. IEEE Transactions on Network and Service Management, 2022, 20(3): 2771-2781.

[41] EL OUADRHIRI A, ABDELHADI A. Differential privacy for deep and federated learning: A survey[J]. IEEE Access, 2022, 10: 22359-22380.

[42] LI Wenling, SUN Jian, JIA Yingmin, et al. Variance-constrained state estimation for nonlinear complex networks with uncertain coupling strength[J]. Digital Signal Processing, 2017, 67: 107-115.

[43] MENG Deyuan, JIA Yingmin, DU Junping. Robust iterative learning protocols for finite-time consensus of multi-agent systems with interval uncertain topologies[J]. International Journal of Systems Science, 2015, 46(5): 857-871.

[44] LI Y W, ZENG I Y, NIU Z H, et al. Predicting vehicle fuel consumption based on multi-view deep neural network[J]. Neurocomputing, 2022, 502: 140-147.

[45] 叶青青, 孟小峰, 朱敏杰, 等. 本地化差分隐私研究综述[J]. 软件学报, 2017, 29(7): 1981-2005.

YE Qingqing, MENG Xiaofeng, ZHU Minjie, et al. A review of localized differential privacy research[J]. Journal of Software, 2017, 29(7): 1981-2005.

[46] BUN M, STEINKE T. Concentrated differential privacy: Simplifications, extensions, and lower bounds[C]//Proceedings of Theory of Cryptography: 14th International Conference, TCC 2016-B. Berlin, Heidelberg: Springer, 2016: 635-658.

[47] TRUEX S, BARACALDO N, ANWAR A, et al. A hybrid approach to privacy-preserving federated learning[C]// Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security. [S.l.]: ACM, 2019: 1-11.

[48] HARDY S, HENECKA W, IVEY-LAW H, et al. Private federated learning on vertically partitioned data via entity resolution and additively homomorphic encryption[EB/OL]. (2017-11-29)[2023-08-01]. https://arxiv.org/abs/1711.10677.

[49] ZHANG Chengliang, LI Suyi, XIA Junzhe, et al. Efficient homomorphic encryption for Cross-Silo federated learning[C]// Proceedings of 2020 USENIX Annual Technical Conference (USENIX ATC 20). [S.l.]: USENIX Association, 2020: 493-506.

[50] FANG Haokun, QIAN Quan. Privacy preserving machine learning with homomorphic encryption and federated learning[J]. Future Internet, 2021, 13(4): 94.

[51] MA J, NAAS S A, SIGG S, et al. Privacy-preserving federated learning based on multi-key homomorphic encryption[J]. International Journal of Intelligent Systems, 2022, 37(9): 5880-5901..

[52] WU Yuncheng, CAI Shaofeng, XIAO Xiaokui, et al. Privacy preserving vertical federated learning for tree-based models[EB/OL]. (2020-08-14)[2023-08-01]. https://arxiv.org/abs/2008.06170.

[53] LI Yawen, DI Jiang, LIAN Rongzhong, et al. Heterogeneous latent topic discovery for semantic text mining[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(1): 533-544

[54] LI Ang, LI Yawen, SHAO Yingxia, et al. Multi-view scholar clustering with dynamic interest tracking[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(9): 9671-9684.

[55] HUANG Jizhou, WANG Haifeng, SUN Yibo, et al. HGAMN: Heterogeneous graph attention matching network for multilingual POI retrieval at baidu maps[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. [S.l.]: ACM, 2021: 3032-3040.

[56] LIM W Y B, LUONG N C, HOANG D T, et al. Federated learning in mobile edge networks: A comprehensive survey[J]. IEEE Communications Surveys & Tutorials, 2020, 22(3): 2031-2063.

[57] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[58] REHMAN M H, DIRIR A M, SALAH K, et al. TrustFed: A framework for fair and trustworthy cross-device federated learning in IIOT[J]. IEEE Transactions on Industrial Informatics, 2021, 17(12): 8485-8494.

[59] 汤文涛. 面向动态聚合的异步联邦学习系统设计与实现[D]. 郑州: 郑州大学, 2022.
TANG Wentao. Design and implementation of an asynchronous federated learning system for dynamic aggregation[D]. Zhengzhou: Zhengzhou University, 2022.

作者简介:

刘星辰(1998-),男,硕士研究生,研究方向:信息检索、数据挖掘和机器学习,E-mail:getlxc@qq.com。

杜军平(1963-),**通信作者**,女,教授,研究方向:人工智能、机器学习、模式识别,E-mail:junpingdu@126.com。

梁美玉(1985-),女,教授,研究方向:人工智能、数据挖掘、多模态数据处理,E-mail: meiyu1210@ bupt. edu.cn。

李昂(1993-),男,博士研究生,研究方向:信息检索、数据挖掘、机器学习,E-mail: david. lee@bupt. edu. cn。

(编辑:刘彦东)