doi:10.3969/j.issn.1672-1292.2025.04.005

基于 PDMM 的联邦 Elastic Net 模型参数安全聚合方案研究

何维民[1],赵　磊[1],余嘉云[2]

(1.国网江苏省电力有限公司营销服务中心,江苏 南京 210019)
(2.南京师范大学组织部(党校),江苏 南京 210023)

[**摘要**]　目前,联邦学习模型均使用数据隐私保护技术(如密码学和差分隐私)来保证模型的参数安全聚合,该技术会带来模型精度低和通信效率低等问题. 为了克服该弊端,本文针对联邦 Elastic Net 模型提出了一种基于原对偶方法(primal-dual method of multipliers,PDMM)的联邦 Elastic Net 参数安全聚合方案——PDMM-Fed. PDMM-Fed 主要分为三步:(1)每个客户端上需要生成一个虚拟客户端,客户端上有训练数据集,虚拟客户端上无训练数据集;(2)将 Elastic Net 的目标函数均方误差项和正则化项分别置于客户端和虚拟客户端上,作为待优化的凸函数;(3)将 PDMM 中的子空间扰动方法引入到中心化的联邦学习网络拓扑中,以确保参与方本地的模型参数不会被逆向推理. 实验结果表明,在保证客户端上模型参数安全的情形下,PDMM-Fed 依然有着较高的通信效率和模型精度.

[**关键词**]　联邦学习模型,Elastic Net,primal-dual method of multipliers,参数安全聚合
[**中图分类号**]TP0812　[**文献标志码**]A　[**文章编号**]1672-1292(2025)04-0037-12

Privacy-Preserving Aggregation Scheme for Federated Elastic Net Model Parameters Based on PDMM

He Weimin[1],Zhao Lei[1],Yu Jiayun[2]

(1.Marketing Service Center,State Grid Jiangsu Electric Power Co. Ltd.,Nanjing 210019,China)
(2.Origanization Department(Party School),Nanjing Normal University,Nanjing 210023,China)

Abstract:At present,data privacy protection technologies (such as cryptography and differential privacy) are all used in Federated Learning model to ensure the secure aggregation of model parameters,which brings problems such as low model accuracy and low communication efficiency. To overcome this drawback,this paper proposes a secure parameter aggregation scheme for the Federated Elastic Net model based on the primal-dual method of multipliers(PDMM),named PDMM-Fed. PDMM-Fed consists of three main steps:(1) Each client generates a virtual client,where the client has a training dataset,while the virtual client does not;(2) The mean squared error term and the regularization term of the Elastic Net objective function are placed on the client and the virtual client,respectively,as convex functions to be optimized;(3) The subspace perturbation method in PDMM is introduced into the centralized Federated Learning network topology to ensure that the local model parameters of participants are not inversely inferred. Experimental results show that PDMM-Fed maintains high communication efficiency and model accuracy while ensuring the security of model parameters on the client side.

Key words:Fedrated Learning model,Elastic Net,primal-dual method of multipliers,parameter secure aggregation

尽管联邦学习模型客户端在全局模型训练中不直接传输本地数据,仅发送模型参数,但由于本地模型参数与数据高度相关,攻击者仍可能通过参数逆向推断出客户端数据,引发隐私泄露. 例如,Fredrikson 等[1]成功地从模型参数中反演出了参与模型训练的图片,且具有一定的辨识度;Melis 等[2]从参与方与服务器通信过程的参数中反演数据,成功地判断出了某个用户的性别.

为了保证模型参数的安全聚合,目前应用在联邦学习参数安全的主流技术主要有 3 种:同态加密

收稿日期:2024-08-06.
基金项目:国网江苏省电力有限公司科技项目(J2020112).
通讯作者:何维民,正高级工程师,研究方向:电力营销数字化. E-mail:13605173535@139.com

(homomorphic encryption)[3-5]、差分隐私(differential privacy)[6]和安全多方计算(secure multiparty computation)(秘密共享(secret sharing)[7-8]). 但是这些技术的的引入,在一定程度上给联邦学习模型带来了模型精度低和通信效率低等问题.

1　相关工作

针对使用同态加密的方法来保护模型参数方法会导致计算和通信开销问题. Zhang 等[9]针对不同组织之间联合建模的场景,提出基于同态加密的安全聚合方案 BatchCrypt,该方案能够提高计算效率,同时减少一定量的通信开销. Mandal 等[10]设计了一种用于联邦环境下对预测模型进行训练和遗忘预测的隐私保护系统 PrivFL,PrivFL 基于加法同态加密方案和聚合协议,可用于训练时的线性和逻辑回归模型的隐私保护.

基于差分隐私的联邦学习参数安全聚合方案可以防止任何第三方还原出原始的模型参数信息. McMahan 等[11]第一次使用差分隐私技术来保证 FedAvg 算法的参数安全聚合. Agarwal 等[12]对二项机制做出了一些改进,并将其应用到分布式向量平均估计(distributed mean estimation)的问题. Sabater 等[13]和 Zhao 等[14]提出了一种差分私有化的平均协议 GOPA(gossip noise for private averaging),该协议可以匹配到可信管理者设置的准确性,同时允许大量的客户端加入联邦学习.

由于联邦学习本质上是多方合作、共同参与计算的一种机器学习,研究者们又将这种特性与安全多方计算进行结合,企图保护模型参数. Kanagavelu 等[15]提出了一种基于安全多方计算的双阶段联邦学习架构. Zhao 等[16]提出一种安全的成员选择策略(secure member selection strategy,SMSS),每个客户端在参与全局模型训练之前需要先评估成员的数据质量.

如上所述,现有的联邦学习参数安全聚合方案存在一些局限性,如需要可信第三方、影响模型精度以及通信代价高等,表 1 对比了 3 种安全聚合方案在通信开销、计算开销和精度等方面的优缺点. 为了打破这些局限,本文针对联邦 Elastic Net 模型,提出了基于原对偶(primal-dual method of multipliers,PDMM)方法的联邦学习参数安全聚合方案(PDMM-Fed),在去中心化的 PDMM 的基础上研究适合 PDMM 优化方法的中心化联邦学习 Elastic Net 模型的网络架构,使用 PDMM 中对偶变量在子空间中不收敛的性质(即子空间扰动法)来保证参与方上的模型参数无法被逆向推理. 实验验证表明,本文所提出的 PDMM-Fed 方案不仅有较高的模型精度和收敛速度,且在联邦学习过程中模型参数一直处于安全的状态. 总体而言,本文所提出的方案具有以下特征:

(1)客户端与服务器之间直接采用明文通信的方式,每轮参数聚合时,服务端与客户端只通信一次;

(2)不需要可信的第三方,不用担心服务端与其他腐败客户端合谋;

(3)只要合适地初始化对偶变量就能保护客户端上的数据隐私,任何人均不能根据客户端上的模型参数准确推断出其本地数据信息;

(4)不同对偶变量的初值对应不同的隐私保护级别,不同的隐私保护级别对模型的收敛率和模型精度不产生影响.

表 1　3 种参数安全聚合方案对比

Table 1　Comparison of three parameter security aggregation schemes

对比项	差分隐私	同态加密	安全多方计算	对比项	差分隐私	同态加密	安全多方计算
通信开销	低	高	高	隐私保护等级	中等	高	高
计算开销	低	高	高	诚实且好奇的服务端	支持	支持	支持
精度	低	高	高	服务器与客户端合谋	支持	不支持	支持

2　预备知识和问题定义

2.1　Elastic Net 模型的目标函数

线性回归的目标函数为:

$$L(\boldsymbol{w})=\frac{1}{2}\|f(\boldsymbol{X};\boldsymbol{w})-\boldsymbol{Y}\|_2^2, \tag{1}$$

式中,$f(\boldsymbol{X};\boldsymbol{w})=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{X}$;$\boldsymbol{w}$ 为模型参数;$\boldsymbol{Y}$ 为所有样本标签组成的列向量;$\boldsymbol{X}$ 为所有样本数据组成的矩阵;$\|\cdot\|_2^2$

为 l_2 范数的平方. 为让线性回归模型具有更好的泛化性和稀疏性,需要在目标函数(1)后添加正则化项,如 l_1 范数、l_2 范数等. 因此,Elastic Net 模型的目标函数即是在线性回归模型目标函数(1)后同时添加 l_1 范数和 l_2 范数的函数. Elastic Net 模型的目标函数为:

$$L_{EN}(\boldsymbol{w})=L(\boldsymbol{w})+R(\boldsymbol{w}), \tag{2}$$

式中,$R(\boldsymbol{w})=r_1\cdot\|\boldsymbol{w}\|_1+r_2\cdot\|\boldsymbol{w}\|_2^2$,$\|\cdot\|_1$ 为 l_1 范数,r_1 和 r_2 为常数.

2.2 联邦学习参数安全聚合问题定义

如图 1 所示,对于客户端 i 上传的模型参数 $\boldsymbol{w}_i$,模型参数采用保护措施生成的函数值记为 $g(\boldsymbol{w}_i)$. 在参数安全保护方案中,客户端 i 在联邦学习的每次全局迭代中发送给服务端的均为 $g(\boldsymbol{w}_i)$. 用互信息来度量第 k 次全局迭代时的隐私保护级别:

$$I(\boldsymbol{w}_i^k;g(\boldsymbol{w}_i^k))=h(\boldsymbol{w}_i)-h(\boldsymbol{w}_i|g(\boldsymbol{w}_i^k)); \tag{3}$$

式中,$\boldsymbol{w}_i^k$ 表示客户端上第 k 次全局迭代的模型参数更新值,$I(\cdot;\cdot)$表示互信息[17]. 当 $I(\boldsymbol{w}_i^k;g(\boldsymbol{w}_i^k))<\varepsilon,\varepsilon>0$ 时,(3)为 ε 统计安全.

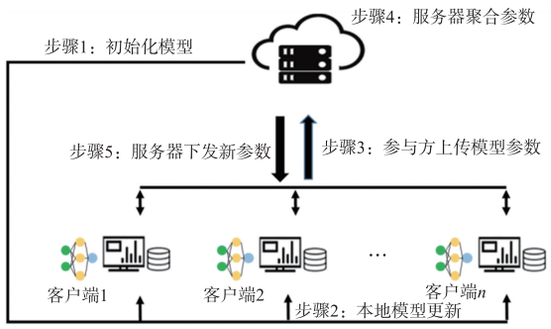

图 1 联邦学习的具体步骤

Fig. 1 Detailed steps of Federated Learning

2.3 原对偶方法

原对偶方法(PDMM)[18-22]是基于 Peaceman-Rachford 分裂的分布式优化理论,具有次线性收敛率. 相比于交替方向乘子法(alternating direction method of multipliers,ADMM),该优化方法包含了更多关于原问题目标函数的信息,因此 PDMM 的收敛速度更快,且参数自身具有广播性质. 文献[23]从单调算子理论的角度对 PDMM 的收敛性和收敛率进行了详细的推导和证明.

PDMM 设计的目标主要是用来解决去中心化的分布式凸优化问题. 在由 m 条边、n 个计算节点构成的网络拓扑图中,将节点 i 上的目标函数记为 $f_i(\boldsymbol{w})$,图的关联矩阵记为 $\boldsymbol{B}$,全局目标函数 $f(\boldsymbol{w})=\sum_{i=1}^{n}f_i(\boldsymbol{w})$.

求解 $f(\boldsymbol{w})$ 的优化问题,可以表示为:

$$\begin{aligned}&\min_{\boldsymbol{w}}\sum_{i=1}^{n}f_i(\boldsymbol{w})\\&\text{s.t. }\boldsymbol{B}_{i|j}\boldsymbol{w}_i+\boldsymbol{B}_{j|i}\boldsymbol{w}_j=0\end{aligned} \tag{4}$$

式中,$\boldsymbol{B}_{i|j}$和 $\boldsymbol{B}_{j|i}$为边(i,j)上的约束矩阵.

使用 PDMM 解决问题(4)时,扩张的增广拉格朗日方程为:

$$\mathcal{L}_\rho(\boldsymbol{w},\boldsymbol{\lambda})=f(\boldsymbol{w})+(\boldsymbol{P}\boldsymbol{\lambda}^r)^{\mathrm{T}}\boldsymbol{C}\boldsymbol{w}+\frac{\rho}{2}\|\boldsymbol{C}\boldsymbol{w}+\boldsymbol{P}\boldsymbol{C}\boldsymbol{w}^r\|_2^2. \tag{5}$$

采用迭代的方式求式(5)的最优解,式(5)中原变量 $\boldsymbol{w}$ 和对偶变量 $\boldsymbol{\lambda}$ 的更新表达式如下所示:

$$\boldsymbol{w}^{r+1}=\arg\min_{\boldsymbol{w}}L(\boldsymbol{w},\boldsymbol{w}^r,\boldsymbol{\lambda}^r), \tag{6}$$

$$\boldsymbol{\lambda}^{r+1}=\rho(\boldsymbol{C}\boldsymbol{w}^{r+1}+\boldsymbol{P}\boldsymbol{C}\boldsymbol{w}^r)-\boldsymbol{\lambda}^r, \tag{7}$$

式中,r 表示迭代次数,$\boldsymbol{w}^r\in\mathbf{R}^u$,$\boldsymbol{\lambda}^r\in\mathbf{R}^{2m}$. $\boldsymbol{P}\in\mathbf{R}^{2m\times 2m}$是正定矩阵,作用是调换矩阵前 m 行和后 m 行. 对于网络拓扑中的任意一条边$(i,j)\in E$ 均包含两个对偶变量 $\boldsymbol{\lambda}_{i|j}\in\mathbf{R}^u$ 和 $\boldsymbol{\lambda}_{j|i}\in\mathbf{R}^u$,$\boldsymbol{\lambda}_{i|j}$在节点 i 上,$\boldsymbol{\lambda}_{j|i}$在节点 j 上,下标 $i|j$ 的标记顺序为$(1|2<1|3<\cdots<1|N<2|1<2|3<\cdots<N|N-1)$. $\rho>0$ 为控制收敛速率的常数,$\boldsymbol{C}=[\boldsymbol{B}_+^{\mathrm{T}}\quad\boldsymbol{B}_-^{\mathrm{T}}]^{\mathrm{T}}$,$\boldsymbol{C}+\boldsymbol{P}\boldsymbol{C}=[\boldsymbol{B}^{\mathrm{T}}\quad\boldsymbol{B}^{\mathrm{T}}]^{\mathrm{T}}$,$\forall(i,j)\in E:\boldsymbol{\lambda}_{i|j}=(\boldsymbol{P}\boldsymbol{\lambda})_{i|j}$.

基于方程(6)和(7),网络拓扑中的每个节点的局部优化变量和对偶变量的更新方程可以表示为:

$$\boldsymbol{w}_i^{r+1}=\arg\min_{\boldsymbol{w}_i}\Big(f_i(\boldsymbol{w}_i)+\sum_{j\in V_i}\boldsymbol{\lambda}_{j|i}^{r,\mathrm{T}}\boldsymbol{B}_{i|j}\boldsymbol{w}_i+\frac{\rho}{2}\sum_{j\in V_i}\|\boldsymbol{B}_{i|j}\boldsymbol{w}_i+\boldsymbol{B}_{j|i}\boldsymbol{w}_j^r\|_2^2\Big), \tag{8}$$

$$\forall j\in V_i:\boldsymbol{\lambda}_{i|j}^{r+1}=\rho(\boldsymbol{B}_{i|j}\boldsymbol{w}_i^{r+1}+\boldsymbol{B}_{j|i}\boldsymbol{w}_j^r)-\boldsymbol{\lambda}_{j|i}^r, \tag{9}$$

式中,V_i 表示节点 i 的邻居节点.

由文献[20-21]可知,式(7)中的对偶变量的连续两次更新有:

$$\boldsymbol{\lambda}^{r+2}=\boldsymbol{\lambda}^{r}+\rho(\boldsymbol{C}\boldsymbol{w}^{r+2}+2\boldsymbol{P}\boldsymbol{C}\boldsymbol{w}^{r+1}+\boldsymbol{C}\boldsymbol{w}^{r}). \tag{10}$$

令 $\overline{\boldsymbol{H}}_p=\mathrm{span}(\boldsymbol{C})+\mathrm{span}(\boldsymbol{PC})$,$\prod_{\overline{\boldsymbol{H}}_p}$为 $\overline{\boldsymbol{H}}_p$ 上的正交投影矩阵,其中 $\mathrm{span}(\boldsymbol{A})$ 表示矩阵 $\boldsymbol{A}$ 列向量张成的空间. 令 $\overline{\boldsymbol{H}}_p^{\perp}=\mathrm{null}(\boldsymbol{C}^{\top})\cap\mathrm{null}((\boldsymbol{PC})^{\top})$,其中 $\mathrm{null}(\boldsymbol{A})$ 表示矩阵 $\boldsymbol{A}^{\mathrm{T}}$ 的核. 由式(10)可知,连续两次 PDMM 对偶变量的更新只影响$\prod_{\overline{\boldsymbol{H}}_p}\boldsymbol{\lambda}\in\overline{\boldsymbol{H}}_p$ 且使得$(\boldsymbol{I}-\prod_{\overline{\boldsymbol{H}}_p})\boldsymbol{\lambda}\in\overline{\boldsymbol{H}}_p^{\perp}$ 不变,$\boldsymbol{I}$ 为单位矩阵. 由式(7)可知,因为 $\boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{I}-\prod_{\overline{\boldsymbol{H}}_p})\boldsymbol{PC}=0$,所以优化变量 $\boldsymbol{w}$ 的更新独立于$(\boldsymbol{I}-\prod_{\overline{\boldsymbol{H}}_p})\boldsymbol{\lambda}$. 因此,$(\boldsymbol{I}-\prod_{\overline{\boldsymbol{H}}_p})\boldsymbol{\lambda}$ 在每次全局迭代过程中只发生行变换,而不收敛. 由文献[22]知,$\prod_{\overline{\boldsymbol{H}}_p}\boldsymbol{\lambda}$ 和$(\boldsymbol{I}-\prod_{\overline{\boldsymbol{H}}_p})\boldsymbol{\lambda}$ 分别为 PDMM 中对偶变量 λ 的收敛和非收敛部分,对于迭代次数 r,可以将对偶变量 $\boldsymbol{\lambda}$ 表示成两个部分加和的形式:

$$\boldsymbol{\lambda}^{r}=\prod_{\overline{\boldsymbol{H}}_p}\boldsymbol{\lambda}^{r}+(\boldsymbol{I}-\prod_{\overline{\boldsymbol{H}}_p})\boldsymbol{\lambda}^{r}\rightarrow\boldsymbol{\lambda}^{*}+\boldsymbol{P}^{r}(\boldsymbol{I}-\prod_{\overline{\boldsymbol{H}}_p})\boldsymbol{\lambda}^{0}. \tag{11}$$

如文献[24]所述,使用 $\boldsymbol{P}^{r}(\boldsymbol{I}-\prod_{\overline{\boldsymbol{H}}_p})\boldsymbol{\lambda}^{0}$ 来保护节点上的数据隐私的方式称为子空间扰动法.

3　基于 PDMM 的联邦 Elastic Net 模型参数安全聚合方案(PDMM-Fed)

3.1　联邦学习网络架构的设计

Elastic Net 模型的目标函数如式(2)所示,该目标函数包含两个部分:均方误差项 $L(\boldsymbol{w})$ 和正则化项 $R(\boldsymbol{w})$. $R(\boldsymbol{w})$可用于控制模型的泛化能力和参数的稀疏性,因此 $R(\boldsymbol{w})$中的系数 r_1,r_2 需要视具体情形设定.

设有 n 个客户端(参与方)参与联邦学习,每个客户端均可与服务器 s 建立通信连接,客户端之间不建立通信连接. 所有客户端上的总的数据集为 $D=(D_1,D_2,\cdots,D_n)$,其中客户端 i 上的数据集为 D_i. 对于任意两个客户端上的数据集 D_i. $D_j(i\neq j)$,$D_i\cap D_j=\phi$,由 D_i 组成的矩阵用 $\boldsymbol{X}_i$ 表示.

客户端 i 上的均方误差函数记为 $L_i(\boldsymbol{w})$,$L_i(\boldsymbol{w})=\frac{1}{2}\|\boldsymbol{X}_i\boldsymbol{w}-\boldsymbol{Y}_i\|_2^2$,$\boldsymbol{Y}_i$ 为数据集矩阵 $\boldsymbol{X}_i$ 对应的标签列向量. 服务端的损失函数 $L_s(\boldsymbol{w})=0$,则基于全样本 D 的均方误差函数 $L(\boldsymbol{w})=\sum_{i=1}^{n}L_i(\boldsymbol{w})+L_s(\boldsymbol{w})=\sum_{i=1}^{n}L_i(\boldsymbol{w})$. 客户端 i 上的正则化函数 $R_i(\boldsymbol{w})$用于控制本地模型的泛化能力与稀疏性. 在所提的联邦学习网络架构中,将客户端 i 上的 $L_i(\boldsymbol{w})$和 $R_i(\boldsymbol{w})$分开,即在客户端 i 分出一个虚拟节点,实结点用于优化 $L_i(\boldsymbol{w})$,虚拟节点用于优化 $R_i(\boldsymbol{w})$[25].

基于上述对联邦学习网络架构的描述,针对 Elastic Net 模型的联邦学习参数安全聚合方案,本文给出如图 2 所示的联邦学习网络架构. 为方便通过变量下标来区分客户端和虚拟客户端,客户端上的变量用 i 表示,虚拟客户端上的变量下标用 i'表示.

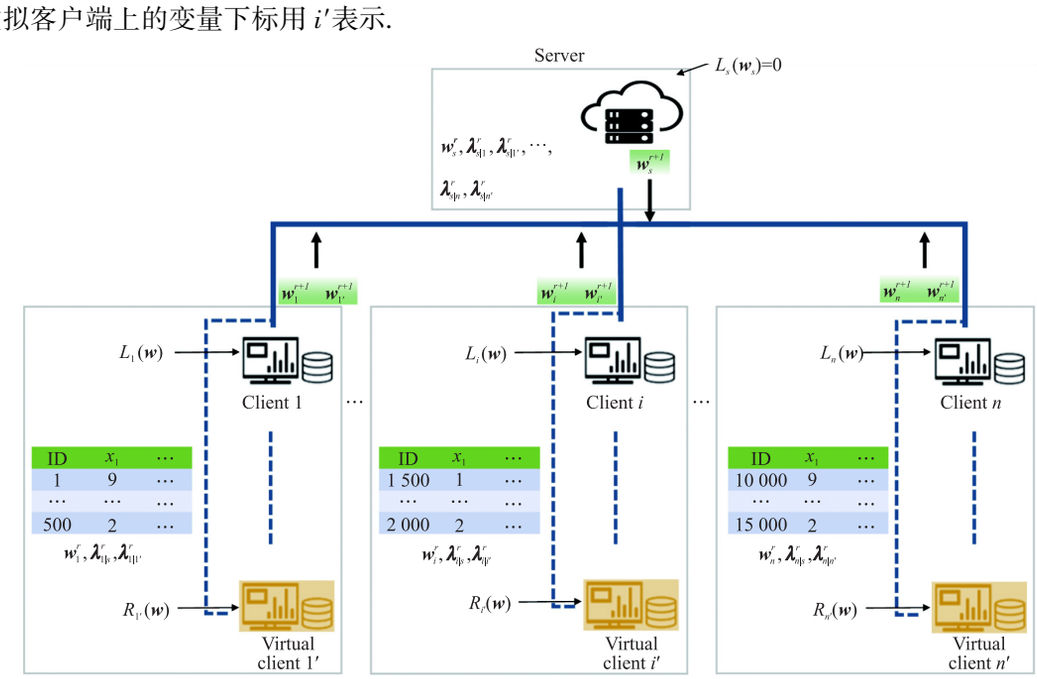

图 2　联邦 Elastic Net 模型参数安全聚合网络架构

Fig. 2　The network architecture of Federated Elastic Net model parameter secure aggregation

3.2 PDMM-Fed 中变量迭代方程以及通信过程

在图 2 所示的网络架构中,对应的联邦学习优化问题可用下式表示:

$$\min_{\{\boldsymbol{w}_s,\boldsymbol{w}_i,\boldsymbol{w}_{i'}\in\mathbf{R}^u\}}\left(\sum_{i=1}^{n}\left[L_i(\boldsymbol{w}_i)+R_{i'}(\boldsymbol{w}_{i'})\right]\right)$$
$$\text{s.t. } \boldsymbol{w}_s=\boldsymbol{w}_i=\boldsymbol{w}_{i'},i=1,\cdots,n. \tag{12}$$

式中,$L_s(\boldsymbol{w})=0$;$\boldsymbol{w}=[\boldsymbol{w}_1,\boldsymbol{w}_{1'},\cdots,\boldsymbol{w}_i,\boldsymbol{w}_{i'},\cdots,\boldsymbol{w}_n,\boldsymbol{w}_{n'},\boldsymbol{w}_s]^{\mathrm{T}}$;$u$ 表示训练样本数据的特征维度. 根据图 2 所示的联邦学习网络架构和式(8)和(9),给出客户端 i、虚拟客户端 i'和服务端 s 上的变量更新表达式. 表达式中对偶变量下标的排列顺序为$(1|s<1'|s<2|s<2'|s<\cdots<n|s<n'|s<s|1<s|1'<\cdots<s|n')$.

客户端 i 上的原变量 $\boldsymbol{w}_i$ 和对偶变量$(\boldsymbol{\lambda}_{i|s},\boldsymbol{\lambda}_{i|i'})$迭代方程为:

$$\begin{cases}\boldsymbol{w}_i^{r+1}=\arg\min\limits_{\boldsymbol{w}_i}\left[L_i(\boldsymbol{w}_i)+(\boldsymbol{\lambda}_{s|i}^r+\boldsymbol{\lambda}_{i'|i}^r)^{\mathrm{T}}\boldsymbol{w}_i+\dfrac{\rho}{2}\|\boldsymbol{w}_i-\boldsymbol{w}_s^r\|^2+\dfrac{\rho}{2}\|\boldsymbol{w}_i-\boldsymbol{w}_{i'}^r\|^2\right],\\ \boldsymbol{\lambda}_{i|s}^{r+1}=\rho(\boldsymbol{w}_s^r-\boldsymbol{w}_i^{r+1})-\boldsymbol{\lambda}_{s|i}^r,\\ \boldsymbol{\lambda}_{i|i'}^{r+1}=\rho(\boldsymbol{w}_{i'}^r-\boldsymbol{w}_i^{r+1})-\boldsymbol{\lambda}_{i'|i}^r.\end{cases} \tag{13}$$

虚拟客户端 i'上的原变量 $\boldsymbol{w}_{i'}$和对偶变量$(\boldsymbol{\lambda}_{i'|s},\boldsymbol{\lambda}_{i'|i})$迭代方程为:

$$\begin{cases}\boldsymbol{w}_{i'}^{r+1}=\arg\min\limits_{\boldsymbol{w}_i}\left[R_{i'}(\boldsymbol{w}_{i'})+(\boldsymbol{\lambda}_{s|i'}^r+\boldsymbol{\lambda}_{i|i'}^r)^{\mathrm{T}}\boldsymbol{w}_{i'}+\dfrac{\rho}{2}\|\boldsymbol{w}_{i'}-\boldsymbol{w}_s^r\|^2+\dfrac{\rho}{2}\|\boldsymbol{w}_{i'}-\boldsymbol{w}_i^r\|^2\right],\\ \boldsymbol{\lambda}_{i'|s}^{r+1}=\rho(\boldsymbol{w}_s^r-\boldsymbol{w}_{i'}^{r+1})-\boldsymbol{\lambda}_{s|i'}^r,\\ \boldsymbol{\lambda}_{i'|i}^{r+1}=\rho(\boldsymbol{w}_i^r-\boldsymbol{w}_{i'}^{r+1})-\boldsymbol{\lambda}_{i|i'}^r.\end{cases} \tag{14}$$

服务端 s 上的原变量 $\boldsymbol{w}_s$ 和对偶变量$\{\boldsymbol{\lambda}_{s|i},\boldsymbol{\lambda}_{s|i'}\}_{i=1,\cdots,n}$的迭代方程为:

$$\begin{cases}\boldsymbol{w}_s^{r+1}=\dfrac{1}{2n}\sum\limits_{i=1}^{n}\left[(\boldsymbol{w}_i^r+\boldsymbol{w}_{i'}^r)-\dfrac{1}{\rho}\sum\limits_{i=1}^{n}(\boldsymbol{\lambda}_{i|s}^r+\boldsymbol{\lambda}_{i'|s}^r)\right],\\ \boldsymbol{\lambda}_{s|i}^{r+1}=\rho(\boldsymbol{w}_i^r-\boldsymbol{w}_s^{r+1})-\boldsymbol{\lambda}_{i|s}^r,\\ \boldsymbol{\lambda}_{s|i'}^{r+1}=\rho(\boldsymbol{w}_{i'}^r-\boldsymbol{w}_s^{r+1})-\boldsymbol{\lambda}_{i'|s}^r.\end{cases} \tag{15}$$

从式(15)易知 $\sum\limits_{i=1}^{n}(\boldsymbol{\lambda}_{s|i}^{r+1}+\boldsymbol{\lambda}_{s|i'}^{r+1})=0$,其中,$\boldsymbol{\lambda}_{i|s},\boldsymbol{\lambda}_{i'|s},\boldsymbol{\lambda}_{s|i},\boldsymbol{\lambda}_{s|i'},\boldsymbol{\lambda}_{i|i'},\boldsymbol{\lambda}_{i'|i}\in\mathbf{R}^u$,$\rho$ 为控制收敛速率的常数.

分析服务端 s、客户端 i 和虚拟客户端 i'在迭代原变量和对偶变量时的网络传输过程:从式(13)-(15)可知,初始情况下,服务端 s 将原变量的初始值 $\boldsymbol{w}_s^0$ 与对偶变量初始值 $\boldsymbol{\lambda}_{s|i}^0$和 $\boldsymbol{\lambda}_{s|i'}^0$一起发送给客户端 i,此时,服务端与客户端之间只建立一次通信连接. 虽然服务端 s 与虚拟客户端 i'有着一条虚拟通信连接,实际上这条通信连接只是逻辑上的,物理上并不存在. 与此同时,客户端 i 将自己的原变量初始值 $\boldsymbol{w}_i^0$、对偶变量初始值 $\boldsymbol{\lambda}_{i|s}^0$和其虚拟客户端 i'上的原变量初始值 $\boldsymbol{w}_{i'}^0$、对偶变量初始值 $\boldsymbol{\lambda}_{i'|s}^0$在一条通信链路上发送给服务端 s. 客户端 i 和虚拟客户端 i'属于同一个物理节点,其间虽有一条逻辑通信连接,但实际上同一个物理节点上的数据是共享的. 所以增加虚拟客户端 i'并未增加通信连接次数. 此后,服务端 s 与客户端 i 之间只需传递原变量的迭代值即可,服务端 s 与客户端 i 和虚拟客户端 i'均可根据上一次迭代完成后的原变量的值来更新最近的对方的对偶变量值. 若服务器不是诚实的,则在图 2 所示的网络架构中,服务器 s 和客户端 i 都有信息:$\{\boldsymbol{w}_s^r,\boldsymbol{\lambda}_{s|i}^r,\boldsymbol{\lambda}_{s|i'}^r\}_{i=1,\cdots,n}\cap\{\boldsymbol{w}_i^r,\boldsymbol{w}_{i'}^r,\boldsymbol{\lambda}_{i|s}^r,\boldsymbol{\lambda}_{i'|s}^r\}_{i=1,\cdots,n}$,客户端 i 和虚拟客户端 i'上的对偶变量$(\boldsymbol{\lambda}_{i|i'}^0,\boldsymbol{\lambda}_{i'|i}^0)$在联邦学习全过程中均未暴露于网络中.

3.3 PDMM-Fed 的参数安全分析

如文献[19-20]所证明的 PDMM 的收敛性,对于任意初始化 $\boldsymbol{w}^0$ 和 $\boldsymbol{\lambda}^0$,原变量 $\boldsymbol{w}^r$ 最终会以收敛率 $O\left(\dfrac{1}{r}\right)$收敛到最优值 $\boldsymbol{w}^*$. 以下同样利用子空间扰动法来分析基于 PDMM 的联邦 Elastic Net 模型训练的参数安全聚合过程.

由于式(13)中原变量 $\boldsymbol{w}_i$ 有闭式解,因此在客户端 i 上有

$$O\in\partial_{\boldsymbol{w}_i}L_i(\boldsymbol{w}_i^{r+1})+\boldsymbol{\lambda}_{s|i}^r+\boldsymbol{\lambda}_{i'|i}^r+\rho(2\boldsymbol{w}_i^{r+1}-\boldsymbol{w}_s^r-\boldsymbol{w}_{i'}^r), \tag{16}$$

式中,$\partial_{\boldsymbol{w}_i}L_i(\boldsymbol{w}_i^{r+1})$为 $L_i(\boldsymbol{w}_i)$在 $\boldsymbol{w}_i^{r+1}$ 上的偏微分. 对于 Elastic Net 模型,易知$\partial_{\boldsymbol{w}_i}L_i(\boldsymbol{w}_i^{r+1})=\boldsymbol{X}_i^{\mathrm{T}}(\boldsymbol{w}_i^{r+1}\boldsymbol{X}_i-\boldsymbol{Y}_i)$. 式

中的原变量 $\boldsymbol{w}_{i'}$有闭式解,因此在虚拟客户端 i'上有

$$O \in \partial_{\boldsymbol{w}_{i'}} R_{i'}(\boldsymbol{w}_{i'}^{r+1}) + \boldsymbol{\lambda}_{s|i'}^{r} + \boldsymbol{\lambda}_{i|i'}^{r} + \rho(2\boldsymbol{w}_{i'}^{r+1} - \boldsymbol{w}_{s}^{r} - \boldsymbol{w}_{i}^{r}), \tag{17}$$

式中,$\partial_{\boldsymbol{w}_{i'}} R_{i'}(\boldsymbol{w}_{i'}^{r+1})$为 $R_{i'}(\boldsymbol{w}_{i'})$在 $\boldsymbol{w}_{i'}^{r+1}$上的偏微分. 对于 Elastic Net 模型,易知

$$\partial_{\boldsymbol{w}_{i'}} R_{i'}(\boldsymbol{w}_{i'}^{r+1}) = r_1 \cdot \operatorname{sign}(\boldsymbol{w}_{i'}^{r+1}) + r_2 \cdot \boldsymbol{w}_{i'}^{r+1}, \tag{18}$$

式中,sign(·)为符号函数,即

$$\operatorname{sign}(\boldsymbol{w}) = \begin{cases} -1, & \boldsymbol{w}_t < 0; \\ 0, & \boldsymbol{w}_t = 0; \\ 1, & \boldsymbol{w}_t > 0. \end{cases} \tag{19}$$

式中,$\boldsymbol{w}_t$ 表示向量 $\boldsymbol{w}$ 中第 t 个分量.

在式(16)中,$\boldsymbol{\lambda}_{s|i}^{r}$归服务器端 s 所有,客户端 i 计算得到 $\boldsymbol{\lambda}_{s|i}^{r}(r>0)$,每个客户端(包括虚拟客户端)的原变量值 $\boldsymbol{w}_i(\boldsymbol{w}_{i'})$暴露于网络中. 因此,对于包括服务器在内的任何第三方,已知的值为:

$$\boldsymbol{\lambda}_{s|i}^{r} + \rho(2\boldsymbol{w}_{i}^{r+1} - \boldsymbol{w}_{s}^{r} - \boldsymbol{w}_{i'}^{r}), \tag{20}$$

而 $\boldsymbol{\lambda}_{i'|i}^{r}$的值属于虚拟参与方 i'上的对偶变量值,该值一直保留在客户端 i 上,未发送给服务端 s. 因此,除了客户端 i 之外,联邦学习网络中的任何计算节点未知的部分为:

$$\partial_{\boldsymbol{w}_i} L_i(\boldsymbol{w}_i^{r+1}) + \boldsymbol{\lambda}_{i'|i}^{r}. \tag{21}$$

类似地,在式(17)中,除客户端 i 之外,联邦学习网络中任何计算节点未知的部分为:

$$\partial_{\boldsymbol{w}_i} R_i(\boldsymbol{w}_i^{r+1}) + \boldsymbol{\lambda}_{i|i'}^{r}. \tag{22}$$

为保证客户端上的数据信息不被泄露,由式(3)知,需要满足

$$I(\partial_{\boldsymbol{w}_i} L_i(\boldsymbol{w}_i^{r+1}); \partial_{\boldsymbol{w}_i} L_i(\boldsymbol{w}_i^{r+1}) + \boldsymbol{\lambda}_{i'|i}^{r}) < \varepsilon. \tag{23}$$

对于图 2 所示的网络拓扑图,当有客户端数量为 n 时,边(包括虚拟边)的数量为 $3n$,则 $\boldsymbol{H} = \operatorname{span}(\boldsymbol{C}_{EN}) + \operatorname{span}(\boldsymbol{P}_{EN}\boldsymbol{C}_{EN})$一定非行满秩. 其中,$\boldsymbol{C}_{EN}$由图 2 所示的网络拓扑图的关联矩阵 $\boldsymbol{B}_{EN}$组成,所以一定有 $\boldsymbol{H}^{\perp} = \operatorname{null}(\boldsymbol{C}_{EN}^{\mathrm{T}}) \cap \operatorname{null}((\boldsymbol{P}\boldsymbol{C}_{EN})^{\mathrm{T}}) \neq 0$ 成立.

通过式(11),可将式(21)等价地表述为:

$$\partial_{\boldsymbol{w}_i} L_i(\boldsymbol{w}_i^{r+1}) + (\prod_{\boldsymbol{H}} \boldsymbol{\lambda}^{r})_{i'|i} + (\boldsymbol{P}^{r}(\boldsymbol{I} - \prod_{\boldsymbol{H}})\boldsymbol{\lambda}^{0})_{i'|i}. \tag{24}$$

式中,$\prod_{\boldsymbol{H}}$ 为矩阵 $\boldsymbol{H}$ 上的正交投影矩阵. 因 $\prod_{\boldsymbol{H}}\boldsymbol{\lambda}^{r}$ 收敛到最优值 $\boldsymbol{\lambda}^{*}$,所以 $\partial_{\boldsymbol{w}_i} L_i(\boldsymbol{w}_i^{r+1}) + (\prod_{\boldsymbol{H}}\boldsymbol{\lambda}^{r})_{i'|i}$的值随着迭代次数有限;又因 $\boldsymbol{H}^{\perp} \neq 0$,当对偶变量的初值 $\boldsymbol{\lambda}^{0} \neq 0$,则 $(\boldsymbol{I} - \prod_{\boldsymbol{H}})\boldsymbol{\lambda}^{0} \neq 0$. 当初始对偶变量时,让 $\boldsymbol{\lambda}^{0}$(方差)足够大,则式(23)一直成立.

结合式(18)和(22),除客户端 i 之外,联邦学习网络中任何计算节点未知的部分为:

$$r_1 \cdot \operatorname{sign}(\boldsymbol{w}_{i'}^{r+1}) + r_2 \cdot \boldsymbol{w}_{i'}^{r+1} + \boldsymbol{\lambda}_{i|i'}^{r}. \tag{25}$$

由式(24)知,只要 $\boldsymbol{\lambda}_{i'|i}^{0}$不泄露,就能保证 $\partial_{\boldsymbol{w}_i} L_i(\boldsymbol{w}_i^{r+1})$不会被正确地推理出来. 但是,由式(13)和(14)易知,在了解 $\boldsymbol{\lambda}_{i|i'}^{r}$时,仍然能很轻松地推理出 $\boldsymbol{\lambda}_{i'|i}^{0}$. 为保证 $\boldsymbol{\lambda}_{i|i'}^{r}$在联邦学习中不被泄露,通过式(11),可将式(22)等价地表述为:

$$r_1 \cdot \operatorname{sign}(\boldsymbol{w}_{i'}^{r+1}) + r_2 \cdot \boldsymbol{w}_{i'}^{r+1} + (\prod_{\boldsymbol{H}} \boldsymbol{\lambda}^{r})_{i|i'} + (\boldsymbol{P}^{r}(\boldsymbol{I} - \prod_{\boldsymbol{H}})\boldsymbol{\lambda}^{0})_{i|i'}. \tag{26}$$

所以当虚拟客户端上的 r_1, r_2 被泄露时,就能够推理出 $\boldsymbol{\lambda}_{i|i'}^{r}$的值,进而可以推理出 $\boldsymbol{\lambda}_{i'|i}^{0}$的值,使得不等式(23)不成立.

由文献[21,23]对互信息中信息泄露的定义,可知当 $(\boldsymbol{P}^{r}(\boldsymbol{I} - \prod_{\boldsymbol{H}})\boldsymbol{\lambda}^{0})_{i'|i}$的方差是 $\partial_{\boldsymbol{w}_i} L_i(\boldsymbol{w}_i^{r+1})$方差的 10 倍时,$I(\partial_{\boldsymbol{w}_i} L_i(\boldsymbol{w}_i^{r+1}); \partial_{\boldsymbol{w}_i} L_i(\boldsymbol{w}_i^{r+1}) + \boldsymbol{\lambda}_{i'|i}^{r}) \approx 0.007$,即 $\partial_{\boldsymbol{w}_i} L_i(\boldsymbol{w}_i^{r+1})$的泄露量只有 0.007 bits. 因此,$\boldsymbol{\lambda}^{0}$ 的方差越大,关于 $\partial_{\boldsymbol{w}_i} L_i(\boldsymbol{w}_i^{r+1})$信息泄露的就越少.

综上所述,这种具有优化与参数安全聚合于一体的联邦 Elastic Net 模型训练方案(PDMM-Fed 方案)的完整算法如下所示.

算法 1　PDMM-Fed 参数安全聚合方案

1:初始化:每个客户端 i 生成一个虚拟客户端 i'. 确定全局迭代次数 R,客户端、虚拟客户端和服务端上的原变量初始值 $\boldsymbol{w}_i^{0}$,$\boldsymbol{w}_{i'}^{0}$,$\boldsymbol{w}_s^{0}$ 可任意赋值,客户端 i 与虚拟客户端 i'之间的对偶变量初值 $\boldsymbol{\lambda}_{i|i'}^{0}$,$\boldsymbol{\lambda}_{i'|i}^{0}$需要足够大,以满足不等式. 客户端 i 上的对

偶变量 $\boldsymbol{\lambda}_{i1s}^0$,虚拟客户端 i' 上的对偶变量 $\boldsymbol{\lambda}_{i'1s}^0$,以及服务端 s 上的对偶变量 $\boldsymbol{\lambda}_{s1i}^0$ 和 $\boldsymbol{\lambda}_{s1i'}^0$ 可以任意初始化,选择合适的 ρ.

2:客户端 i 将 $\boldsymbol{w}_i^0,\boldsymbol{w}_{i'}^0,\boldsymbol{\lambda}_{i1s}^0,\boldsymbol{\lambda}_{s1i'}^0$ 发送给服务端 s,服务端 s 将 $\boldsymbol{w}_s^0,\boldsymbol{\lambda}_{s1i}^0$ 和 $\boldsymbol{\lambda}_{s1i'}^0$ 发送给客户端 i,客户端 i 和虚拟客户端 i' 共享所有变量值.

3:For $r=0,\cdots,R$ do:

4: 服务端 s:根据式(15)可得 $\boldsymbol{w}_s^{r+1},\boldsymbol{\lambda}_{s1i}^{r+1},\boldsymbol{\lambda}_{s1i'}^{r+1}$,并将 $\boldsymbol{w}_s^{r+1}$ 发送给客户端.

5: 客户端 i 和虚拟客户端 i':

6: 原变量的更新:

根据式(13)和(14)可得 $\boldsymbol{w}_i^{r+1},\boldsymbol{w}_{i'}^{r+1}$.

7: 对偶变量的更新:

根据式(13)和(14)可得 $\boldsymbol{\lambda}_{i1s}^{r+1},\boldsymbol{\lambda}_{i1i'}^{r+1},\boldsymbol{\lambda}_{i'1s}^{r+1},\boldsymbol{\lambda}_{i'1i}^{r+1}$.

8: 客户端 i 将 $\boldsymbol{w}_i^{r+1},\boldsymbol{w}_{i'}^{r+1}$ 发送给服务端 s.

9: 当客户端接收到服务端 s 的 $\boldsymbol{w}_s^{r+1}$ 时,客户端 i 更新服务端 s 的对偶变量:

根据式(15)可得 $\boldsymbol{\lambda}_{s1i}^{r+1},\boldsymbol{\lambda}_{s1i'}^{r+1}$.

10: 当服务端 s 接收到 $\boldsymbol{w}_i^{r+1},\boldsymbol{w}_{i'}^{r+1}$ 时,服务端 s 更新客户端 i 和虚拟客户端 i' 上的对偶变量:

根据式(13)可得 $\boldsymbol{\lambda}_{i1s}^{r+1}$,根据式(14)可得 $\boldsymbol{\lambda}_{i'1s}^{r+1}$.

11: If $\|\boldsymbol{w}_i^r-\boldsymbol{w}_{i'}^r\|_2^2\leqslant$threshold and $\|\boldsymbol{w}_i^r-\boldsymbol{w}_s^r\|_2^2\leqslant$threshold:End for

12:End for

4 实验与分析

为验证本文研究方案的效果,采用大小两个数据集进行实验验证,小数据集采用 Boston 房价数据集,大数据集采用 California 房价数据集,这两个数据集均为线性回归常用公开数据集. 通过对比 PDMM-Fed 方案与 FedAvg、SCAFFOLD 和 FedProx 算法验证本文算法在性能上的优化. 为验证式(24)的正确性,在全局迭代过程中,保持对偶变量的非收敛部分恒定,从而检验本文算法在参数安全聚合方面的性能.

4.1 Boston 房价数据集上的实验与分析

4.1.1 优化性能对比实验

由于 Boston 房价数据集较小,实验设置客户端数量为 $n\in\{2,4,8\}$. 每个客户端上的数据集之间均是 Non-IID 的. PDMM-Fed 中的常数 $\rho=10$,SCAFFOLD 算法中的 $\mu_g=1$,FedProx 中的 $\mu=1$. 在 4 种优化方案中,模型初始化参数 $\boldsymbol{w}_i^0,\boldsymbol{w}_s^0$ 均为 0,其中 PDMM-Fed 方案中客户端与虚拟客户端之间的对偶变量 $\boldsymbol{\lambda}_{i1i'}^0$、$\boldsymbol{\lambda}_{i'1i}^0$ 采用满足高斯分布(均值为 0,方差为 1×10^4)的初始化,其他的对偶变量均初始化为 0,梯度步长均为 1×10^{-4}. FedAvg、SCAFFOLD 和 FedProx 算法采用了每个客户端本地进行 20 次局部梯度下降的方法,每个客户端上的正则化系数 $r_1,r_2\in(0,1)$. PDMM-Fed 方案中,正则化系数 $r_1,r_2\in(0,1)$ 且每个虚拟客户端上的正则化系数均不同.

图 3 记录了不同联邦学习优化算法的性能. 图 3 的每个子图横轴表示联邦学习的全局迭代次数,纵轴表示每轮全局迭代对应的损失函数值,其中左侧纵轴表示 FedAvg、SCAFFOLD 和 FedProx 的全局损失函

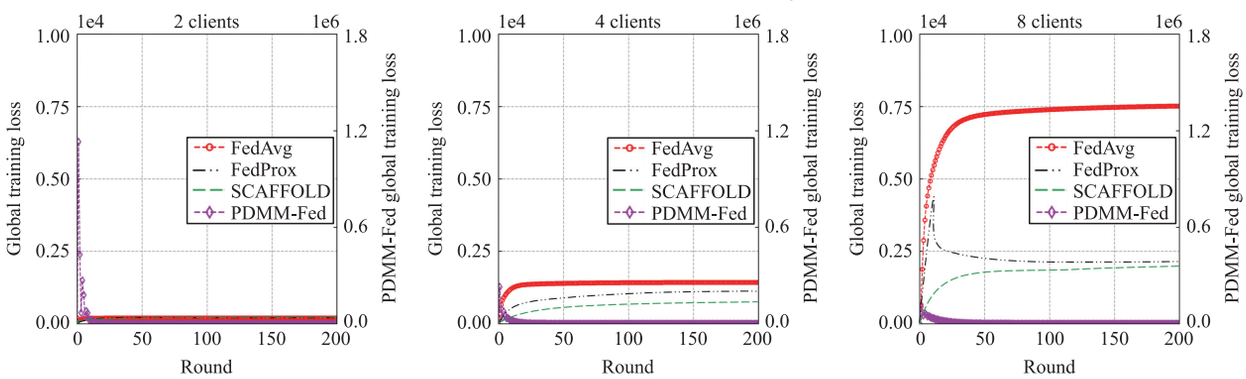

图 3 Boston 房价数据集上 FedAvg、SCAFFOLD、FedProx 和 PDMM-Fed 优化性能对比

Fig. 3 Comparison of optimization performance of FedAvg, SCAFFOLD, FedProx, and PDMM-Fed on the Boston housing dataset

数值,右侧纵轴表示 PDMM-Fed 全局损失函数值.

评价优化算法的性能主要从收敛速率和收敛值两个方面考虑. 在收敛速率方面,PDMM-Fed 方案不亚于其他 3 种优化算法. 随着客户端数量的增加,4 种联邦学习优化算法的收敛速率均有所下降. 在收敛值方面,PDMM-Fed 方案优于其他 3 种联邦学习优化算法. 随着客户端数量的增加,FedAvg、SCAFFOLD 和 FedProx 算法的收敛值均有明显增加,而 PDMM-Fed 算法的收敛值无明显增加或增加得较小.

表 2 所示为 FedAvg、SCAFFOLD、FedProx 和 PDMM-Fed 算法在 Bonston 房价测试数据集上的精度,使用 Mean squared error(MSE)和 R square(R^2)来度量模型精度,其中,MSE 越小表示模型精度越高,R^2 值越大表示模型精度越高. 从表 2 可以看出,无论客户端数量是多少,PDMM-Fed 方案训练出的模型精度均优于其他 3 种优化算法. 随着客户端数量的增加,4 种优化算法训练出的模型精度均有所降低,与 FedAvg、SCAFFOLF 和 FedProx 算法不同的是,PDMM-Fed 训练出的模型精度受客户端数量影响较小,受客户端数量影响最大的是 FedAvg 算法.

表 2　Boston 数据集上 4 种联邦学习优化算法的模型精度对比

Table 2　Comparison of model accuracy of four Federated Learning optimization algorithms on the Boston dataset

指标	算法	2 Clients	4 Clients	8 Clients	指标	算法	2 Clients	4 Clients	8 Clients
MSE	FedAvg	1.191	7.511	38.128	R^2	FedAvg	0.608	0.385	0.214
	SCAFFOLD	0.926	1.647	5.678		SCAFFOLD	0.612	0.556	0.571
	FedProx	0.933	0.534	5.799		FedProx	0.620	0.533	0.568
	PDMM-Fed	0.369	0.364	0.362		PDMM-Fed	0.639	0.624	0.594

4.1.2　参数安全验证

图 4 记录了 Boston 房价数据集上从不同客户端数量的 PDMM-Fed 方案中随机选取一个客户端 i 上的对偶变量的迭代情况,其中任意一个客户端 i 上的对偶变量 $\boldsymbol{\lambda}^0_{i'|i}$ 的方差为 1×10^4. 图 4 的横轴表示全局迭代次数,左侧纵轴表示对偶变量收敛部分的方差,右侧纵轴表示对偶变量非收敛部分的方差. 从图 4 中可以看出,不同客户端数量的 PDMM-Fed 方案对偶变量的收敛部分在后续全局迭代过程中保持恒定. 图 4 的实验结果即可验证式(24)的正确性.

针对不同的参数安全保护级别,可对客户端 i 上的对偶变量初值 $\boldsymbol{\lambda}^0_{i'|i}$ 和 $\boldsymbol{\lambda}^0_{i|i'}$ 采用不同的初始化策略. 当隐私保护级别越高,$\boldsymbol{\lambda}^0_{i'|i}$ 的方差就需要越大. 图 5 记录了不同隐私保护级别下的对偶变量 $\boldsymbol{\lambda}$ 的非收敛部分 $((\boldsymbol{I}-\prod_{\boldsymbol{H}})\boldsymbol{\lambda}^0)_{i'|i}$ 在全局迭代过程中的迭代情况,其中横轴表示全局迭代次数,纵轴表示 $((\boldsymbol{I}-\prod_{\boldsymbol{H}})\boldsymbol{\lambda}^r)_{i'|i}$ 的方差. 从图 5 可以看出,当 $\boldsymbol{\lambda}^0_{i'|i}$ 的方差越大,对应的 $((\boldsymbol{I}-\prod_{\boldsymbol{H}})\boldsymbol{\lambda}^r)_{i'|i}$ 方差就越大,隐私保护级别就越高.

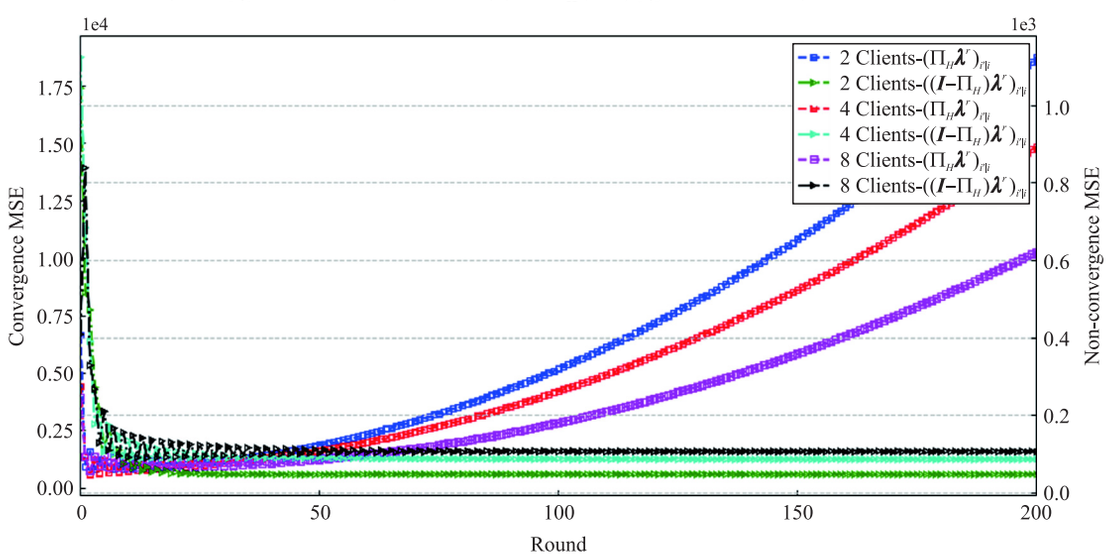

图 4　客户端 i 上对偶变量的收敛部分和非收敛部分迭代情况

Fig. 4　Convergence and non-convergence parts of dual variables iteration on client i

4.2 California 数据集上的实验与分析

4.2.1 优化性能对比实验

在 California 房价数据集上,设置客户端数量为 $n\in\{2,5,10,20\}$,其他设置与 Boston 房价数据集实验设置相同.

图 6 所示为不同联邦学习优化算法的性能. 图 6 的每个子图横轴表示联邦学习的全局迭代次数,纵轴表示每轮全局迭代对应的损失函数值,其中左侧纵轴表示 FedAvg、SCAFFOLD 和 FedProx 的全局损失函数值,右侧纵轴表示 PDMM-Fed 全局损失函数值.

表 3 所示为 FedAvg、SCAFFOLD、FedProx 和 PDMM-Fed 算法在 California 房价测试数据集上的精度. 从表 3 可以看出,对于绝大多数情况,PDMM-Fed 方案训练出的模型精度均优于其他 3 种优化算法. 随着客户端数量的增加,4 种优化算法训练出的模型精度均有所降低,与 FedAvg、SCAFFOLF 和 FedProx 算法不同的是,PDMM-Fed 方案训练出的模型精度受客户端数量影响较小,受客户端数量影响最大的是 FedAvg 算法.

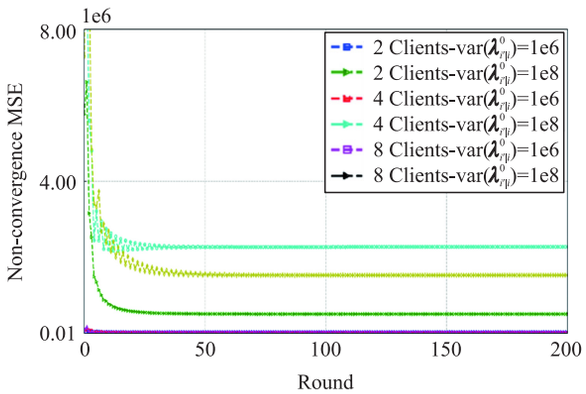

图 5 Boston 数据集上不同隐私保护级别下的对偶变量非收敛部分迭代图

Fig. 5 Dual variable non-convergence part iteration diagram under different privacy protection levels on the Boston dataset

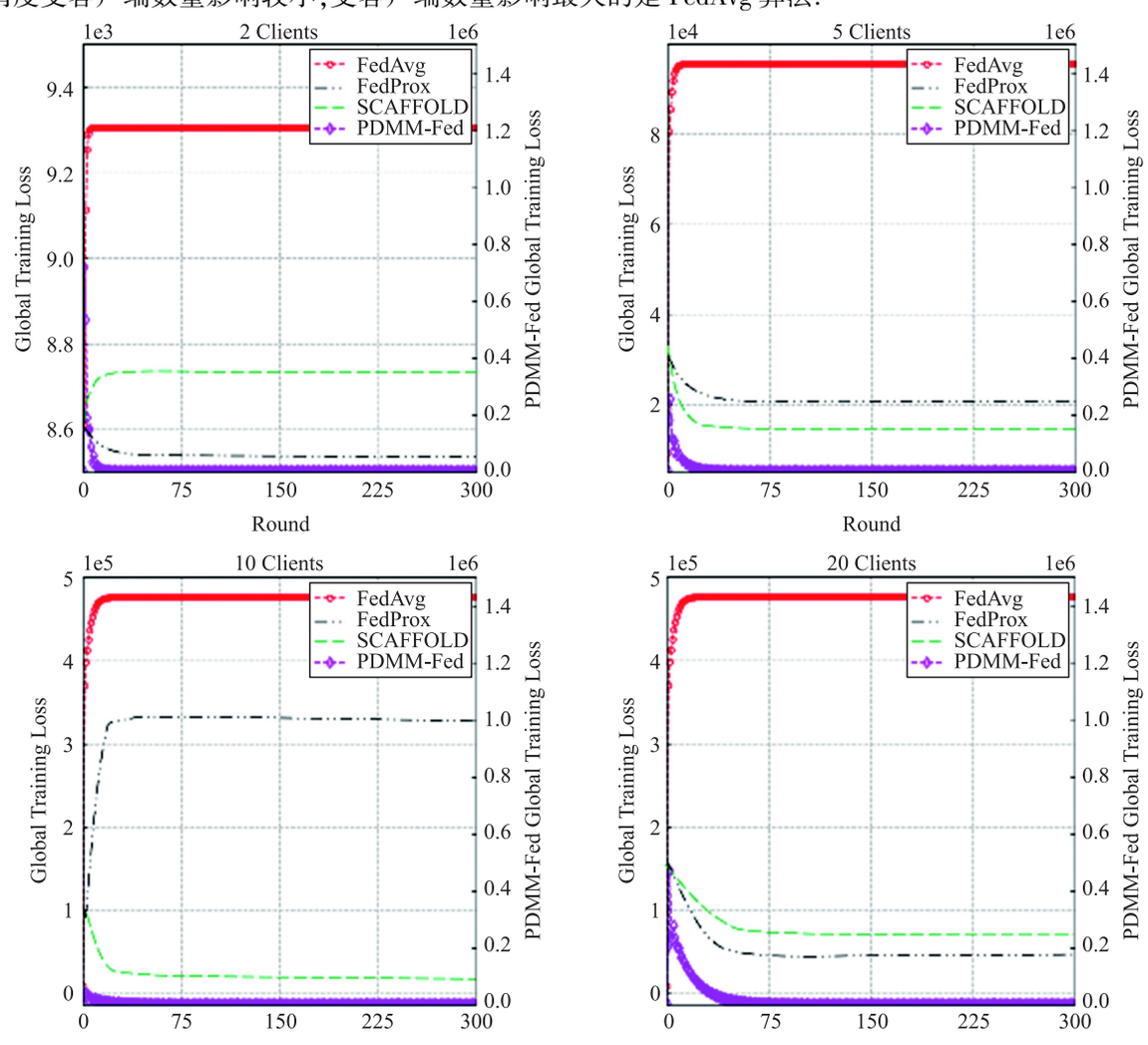

图 6 California 数据集上 FedAvg、SCAFFOLD、FedProx 和 PDMM-Fed 优化性能对比图

Fig. 6 Comparison of optimization performance of FedAvg, SCAFFOLD, FedProx, and PDMM-Fed on the California dataset

表 3　California 数据集上 4 种联邦学习优化算法的模型精度

Table 3　Model accuracy of four Federated Learning optimization algorithms on the California dataset

指标	算法	2 Clients	5 Clients	10 Clients	20 Clients	指标	算法	2 Clients	5 Clients	10 Clients	20 Clients
MSE	FedAvg	0.970	10.080	50.850	50.854	R^2	FedAvg	0.590	0.337	0.185	0.181
	SCAFFOLD	0.638	2.064	4.567	10.335		SCAFFOLD	0.545	0.345	0.215	0.266
	FedProx	0.545	2.789	5.679	6.986		FedProx	0.513	0.422	0.399	0.405
	PDMM-Fed	0.405	0.404	0.402	0.403		PDMM-Fed	0.500	0.497	0.495	0.494

4.2.2　隐私保护验证

图 7 所示为 California 房价数据集上从不同客户端数量的 PDMM-Fed 方案中随机选取一个客户端 i 上的对偶变量的迭代情况,其中任意一个客户端 i 上的对偶变量 $\boldsymbol{\lambda}^0_{i'|i}$ 的方差为 1×10^6. 图 7 的横轴表示全局迭代次数,左侧纵轴表示对偶变量收敛部分的方差,右侧纵轴表示对偶变量非收敛部分的方差. 从图 7 可以看出,不同客户端数量的 PDMM-Fed 方案对偶变量的非收敛部分在后续全局迭代过程中保持恒定. 图 7 的实验结果即可验证式(24)的正确性.

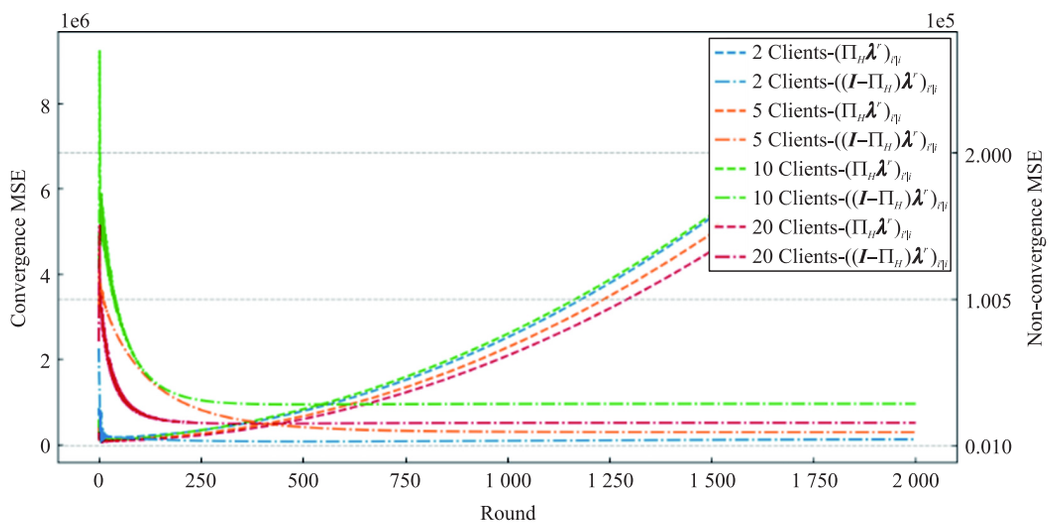

图 7　客户端 i 上对偶变量的收敛部分和非收敛部分迭代图

Fig. 7　Convergence and non-convergence parts of dual variables iteration on client i

图 8 所示为 california 房价数据集上的 PDMM-Fed 方案中不同隐私保护级别下的对偶变量 $\boldsymbol{\lambda}$ 的非收敛部分$((\boldsymbol{I}-\prod_{\boldsymbol{H}})\boldsymbol{\lambda}^0)_{i'|i}$在全局迭代过程中的迭代情况,其中横轴表示全局迭代次数,纵轴表示$((\boldsymbol{I}-\prod_{\boldsymbol{H}})\boldsymbol{\lambda}^r)_{i'|i}$的方差. 从图 8 可以看出,当 $\boldsymbol{\lambda}^0_{i'|i}$ 的方差越大,对应的$((\boldsymbol{I}-\prod_{\boldsymbol{H}})\boldsymbol{\lambda}^r)_{i'|i}$方差就越大,隐私保护级别就越高.

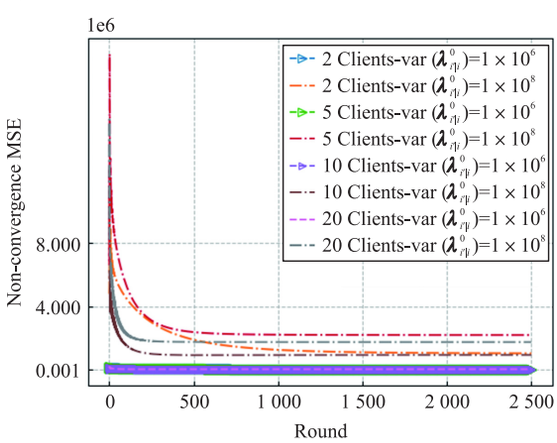

图 8　California 数据集上不同隐私保护级别下的对偶变量非收敛部分迭代图

Fig. 8　Iteration graph of non-convergence parts of dual variables under different privacy protection levels on the California dataset

5　结论

目前的联邦学习参数安全聚合方案均采用数据隐私保护技术,暂无直接从优化算法入手的参数安全聚合方案. 本文提出了基于 PDMM 的联邦 Elastic Net 模型参数安全聚合方案,是一种 PDMM 算法在中心化联邦学习上的应用,并使用 PDMM 算法中的子空间扰动性质来保证联邦学习中的参数安全. 从实验的结果可知,在非独立同分布的样本上 PDMM-Fed 方案在优化性能上优于目前的联邦优化算法,同时验证了参数安全的可行性. PDMM-Fed 方案是将 PDMM 应用于联邦机器学习模型全局优化和参数安全聚合一次尝试,将来可将 PDMM-Fed 方案的思想应用于更多的机器学习模型.

[参考文献](References)

[1] FREDRIKSON M, JHA S, RISTENPART T. Model inversion attacks that exploit confidence information and basic countermeasures[C]//Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. Denver,USA:ACM,2015.

[2] MELIS L, SONG C Z, DE CRISTOFARO E, et al. Exploiting unintended feature leakage in collaborative learning[C]//Proceedings of the 2019 IEEE Symposium on Security and Privacy(SP). San Francisco,USA:IEEE,2019.

[3] GENTRY C. Fully homomorphic encryption using ideal lattices[C]//Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing. Bethesda,USA:ACM,2009.

[4] CORON J S,LEPOINT T,TIBOUCHI M. Scale-Invariant fully homomorphic encryption over the integers[M]//KRAWCZYK H. Public-Key Cryptography-PKC 2014. Berlin,Germany:Springer,2014.

[5] HE W M,ZHAO L,CHENG L F. A homomorphic encryption method for power data based on improved paillier algorithm[M]//SUN X M,ZHANG, X R,XIA Z H,et al. Advances in Artificial Intelligence and Security. Switzerland:Springer Cham,2021.

[6] DWORK C. Differential privacy: A survey of results[M]//AGRAWAL M, DU D Z, DUAN Z H, et al. Theory and Applications of Models of Computation. Berlin,Germany:Springer,2008.

[7] CHANDRAMOULI A,CHOUDHURY A,PATRA A. A survey on perfectly secure verifiable secret-sharing[J]. ACM Computing Surveys(CSUR),2022,54(11s):232.

[8] GUO C, HANNUN A, KNOTT B, et al. Secure multiparty computations in floating-point arithmetic[J]. Information and Inference:A Journal of the IMA,2022,11(1):103-135.

[9] ZHANG C L,LI S Y,XIA J Z,et al. BatchCrypt:Efficient Homomorphic Encryption for Cross-Silo Federated Learning[C]//Proceedings of the 2020 USENIX Annual Technical Conference(USENIX ATC 2020). Online:USENIX,2020.

[10] MANDAL K,GONG G. PrivFL:Practical privacy-preserving federated regressions on high-dimensional data over mobile networks[C]//Proceedings of the 2019 ACM SIGSAC Conference on Cloud Computing Security Workshop. London,UK:ACM,2019.

[11] MCMAHAN H B,RAMAGE D,TALWAR K,et al. Learning differentially private recurrent language models[C]//Proceedings of the 35th International Conference on Learning Representations. Stockholm,Sweden:ICLR,2018.

[12] AGARWAL N,SURESH A T,YU F X,et al. cpSGD:Communication-efficient and differentially-private distributed SGD[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montreal,Canada:NIPS,2018.

[13] SABATER C, BELLET A, RAMON J. Distributed differentially private averaging with improved utility and robustness to malicious parties[EB/OL]. (2020-06-12) [2024-08-06]. https://doi.org/10.48550/arXiv.2006.07218.

[14] ZHAO Y,ZHAO J,YANG M M,et al. Local differential privacy-based federated learning for internet of things[J]. IEEE Internet of Things Journal,2021,8(11):8836-8853.

[15] KANAGAVELU R,LI Z,SAMSUDIN J,et al. Two-phase multi-party computation enabled privacy-preserving federated learning[C]//Proceedings of the 20th IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing(CCGRID). Melbourne,Australia:IEEE,2020.

[16] ZHAO K,XI W,WANG Z,et al. SMSS:Secure member selection strategy in federated learning[J]. IEEE Intelligent Systems,2020,35(4):37-49.

[17] COVER T M. Elements of Information Theory[M]. New Jersey,USA:John Wiley & Sons,1999.

[18] ZHANG G Q,HEUSDENS R. Bi-alternating direction method of multipliers[C]//Proceedings of the 2013 IEEE International Conference on Acoustics,Speech and Signal Processing. Vancouver,Canada:IEEE,2013.

[19] ZHANG G Q,HEUSDENS R. Bi-alternating direction method of multipliers over graphs[C]//Proceedings of the 2015 IEEE International Conference on Acoustics,Speech and Signal Processing(ICASSP). South Brisbane,Australia:IEEE,2015.

[20] WANG J T ,WANG Y C. Designing unimodular sequences with optimized auto/cross-correlation properties via consensus-ADMM/PDMM approaches[J]. IEEE Transactions on Signal Processing,2021,69:2987-2999.

[21] ZHANG G Q,HEUSDENS R. Distributed optimization using the primal-dual method of multipliers[J]. IEEE Transactions on Signal and Information Processing over Networks,2017,4(1):173-187.

[22] 徐占洋,程洛飞,程建春,等. 一种使用 Bi-ADMM 优化深度学习模型的方案[J]. 信息网络安全,2023,23(2):54-63.

[23] SHERSON T W,HEUSDENS R,KLEIJN W B. Derivation and analysis of the primal-dual method of multipliers based on monotone operator theory[J]. IEEE Transactions on Signal and Information Processing over Networks,2018,5(2):334-347.

[24] LI Q X,HEUSDENS R,CHRISTENSEN M G. Privacy-preserving distributed optimization via subspace perturbation:A general

framework[J]. IEEE Transactions on Signal Processing,2020,68:5983-5996.
[25] O'CONNOR M,ZHANG G Q,KLEIJN W B,et al. Function splitting and quadratic approximation of the primal-dual method of multipliers for distributed optimization over graphs[J]. IEEE Transactions on Signal and Information Processing over Networks,2018,4(4):656-666.

[责任编辑:严海琳]

(上接第 27 页)

[26] CHURCH K W. Word2Vec[J]. Natural Language Engineering,2017,23(1):155-162.
[27] 谢玉惠,肖桂荣. 融合注意力机制的多通道 CNNs-BiLSTM 情感极性分析方法[J]. 小型微型计算机系统,2023,44(6):1140-1145.
[28] ZHANG H W,WANG J,ZHANG J X,et al. YNU-HPCC at SemEval 2017 task 4:Using a multi-channel CNN-LSTM model for sentiment classification[C]//Proceedings of the 11th International Workshop on Semantic Evaluation(SemEval-2017). Vancouver,Canada:ACL,2017.
[29] 程艳,尧磊波,张光河,等. 基于注意力机制的多通道 CNN 和 BiGRU 的文本情感倾向性分析[J]. 计算机研究与发展,2020,57(12):2583-2595.
[30] BI X,ZHANG T. Pedagogical sentiment analysis based on the BERT-CNN-BiGRU-attention model in the context of intercultural communication barriers[J]. PeerJ Computer Science,2024,10:e2166.
[31] VANGUMALLA R S,CHOI Y. Exploring sentiment analysis:A study on rheumatoid arthritis and lupus in healthcare[J]. Research Reports on Computer Science,2024:34-60-34-60.

[责任编辑:严海琳]