doi:10.3969/j.issn.1001-4616.2024.04.001

时空数据隐私保护共享的群体学习方法研究

康海燕,吴思远

(北京信息科技大学计算机学院,北京 100192)

[**摘要**] 实现时空数据的共享流通及协同分析能够挖掘数据潜在价值、助力地理信息产业发展,但私有数据的隐私泄露抑制了时空数据的共享. 为了在进一步推动时空数据共享程度、优化共享效果的同时兼顾参与方经济效益及合法权益,提出了一种时空数据隐私保护共享的群体学习(spatio-temporal data privacy preserving sharing swarm learning,STDPPS-SL)方法. 首先,构建基于群体学习的多参与方时空数据共享网络,保护参与方数据所有权,实现数据内容不泄漏且参与方权益平等的时空数据共享;其次,提出基于 t 分布的差分隐私随机梯度下降算法,防止共享过程中因隐私泄露导致参与方私有数据保密性被破坏,从而造成参与方经济损失;最后,设计打分系统量化参与方可信程度,保证数据共享结果可信. 理论分析证明,本文所提方法(STDPPS-SL)满足严格差分隐私,能够保护参与方的私有数据所有权. 在公开数据集上的对比实验表明,该方法(STDPPS-SL)能够实现参与方隐私保护的时空数据共享,并且兼顾安全性与可用性.

[**关键词**] 数据共享,时空大数据,群体学习,分布式学习

[**中图分类号**]TP181　[**文献标志码**]A　[**文章编号**]1001-4616(2024)04-0001-10

Research on Spatio-Temporal Data Privacy Preserving Sharing Swarm Learning

Kang Haiyan,Wu Siyuan

(Computer School,Beijing Information Science and Technology University,Beijing 100192,China)

Abstract:Achieving the sharing and collaborative analysis of spatio-temporal data can explore the potential value of data and boost the development of geographic information industry,but the privacy leakage of private data inhibits the sharing of spatio-temporal data.In order to further promote the degree of spatio-temporal data sharing and optimize the sharing effect while taking into account the economic benefits and legitimate rights of the participants,this paper proposes a spatio-temporal data privacy preserving sharing swarm learning(STDPPS-SL) method. Firstly,a multi-participant spatio-temporal data sharing network based on swarm learning is proposed in order to protect the ownership of participant's data and enable the process of spatio-temporal data sharing without revealing the contents of the data,while keep the equal rights of participants. Secondly,a differential privacy stochastic gradient descent algorithm based on the t-distribution is proposed in order to prevent the confidentiality of the participant's private data from being destroyed due to the privacy leakage during the sharing process,and in order to avoid economic losses to the participants. Finally,a scoring system is designed to quantify the credibility of the participants,in order to ensure the credibility of the data sharing results. Theoretical analysis proves that the proposed method(STDPPS-SL)can protect the private data ownership of the participants by satisfying strict differential privacy. Comparative experiments on open datasets show that the proposed method(STDPPS-SL)is able to realize the spatio-temporal data sharing process with the protection of participant's privacy,and the method balances the security and usability.

Key words:data sharing,spatio-temporal big data,swarm learning,distributed learning

收稿日期:2024-05-16.

基金项目:国家社会科学基金项目(21BTQ079)、教育部人文社科项目(20YJAZH046)、未来区块链与隐私计算高精尖创新中心基金项目(GJJ-23).

通讯作者:康海燕,博士,教授,研究方向:网络安全与隐私计算. E-mail:kanghaiyan@126.com

时空数据是地理信息行业发展的基础资源和关键要素,实现海量时空数据的共享流通、挖掘利用其内在价值已成为地理信息相关产业在大数据时代背景下的发展机遇. 与此同时,国家加大数据合规监管力度,相继出台多部数据安全法律. 在时空数据共享需求激增、数据安全合规日趋收紧的背景下,实现数据安全、隐私保护的时空数据共享已成为我国地理信息产业拥抱大数据时代的关键方式.

与直接交换私有数据内容的传统时空数据共享方法[1-3]相比,联邦学习[4-6]通过共享各参与方使用私有数据训练的本地模型参数,实现“数据可用不可见”“数据不动模型动”的时空数据共享,是大数据时代背景下地理信息行业实现时空数据共享的更优选择. 基于上述背景,关于联邦学习的时空数据共享方案相关研究层出不穷. Graser 等[7]指出,即便当前地理信息系统科学界对联邦学习的研究力度日趋增大,但联邦学习在时空数据共享、地理分析和空间统计等领域依旧存在广阔的研究空间;Ashish[8]提出了一种用户时空数据共享框架,通过将联邦学习与地理空间语义分析相结合,在显著提高用户隐私保护水平的同时,保持基于位置的服务(location-based services,LBS)的效果稳定可靠;Chung 等[9]提出了一种基于联邦学习的用户下一步位置预测方案 FedGeo,使用联邦学习实现用户地理位置时空数据共享,并通过接入先验的全球地理邻接信息、引入全新聚合方法、排除极度异构数据保证预测准确性;蒋伟进等[10]提出了一种基于自适应联邦学习的环境监测数据共享方法,通过结合联邦学习与边缘计算,构建可扩展的时空数据共享架构,并采用动态优化迭代频率策略,优化数据共享效果.

基于上述内容,可知现有的基于联邦学习的时空数据共享方法在诸多地理信息相关研究领域均发挥作用. 然而,现有的联邦学习时空数据共享方案存在数据共享效果及参与方权益问题,主要表现在三方面:首先,联邦学习通过聚合各参与方本地模型实现无需传输参与方私有数据的数据共享,为了获取更好的数据共享效果,有必要分辨不可信的参与方,并调整对不同参与方本地模型的聚合权重. 其次,联邦学习在网络中设计了固定的中心服务器,其能够获取网络中任意参与方的本地模型参数,并负责聚合各方本地模型以生成网络全局模型. 若恶意参与方攻击并控制中心服务器,将威胁网络中各参与节点的合法权益. 最后,不传输数据、仅传输模型参数能够实现数据共享过程的隐私保护是现有基于联邦学习的时空数据共享研究领域的共识. 但最新研究表明,存在针对模型参数的分析方法[11-13]使得攻击者能够基于模型参数重建参与方私有本地时空数据,单一的联邦学习存在隐私泄露风险.

针对上述问题,本文进行了深入研究,主要贡献如下:

① 提出一种时空数据隐私保护共享的群体学习(spatio-temporal data privacy preserving sharing swarm learning,STDPPS-SL)方法,实现无需传输数据的时空数据隐私保护共享及协同分析.

② 提出基于 t 分布的差分隐私随机梯度下降(t-differential privacy stochastic gradient descent,t-dpSGD)算法. 在传统随机梯度下降的基础上对模型参数进行梯度裁剪和 t 分布随机噪声扰动,实现参与方模型参数训练及本地化差分隐私[14]处理,解决时空数据共享过程中的敏感信息隐私泄露和数据重构问题.

③ 设计参与方打分系统,量化参与方可信程度,结合参与方分数设计聚合节点选择算法及自适应模型聚合算法,选择可靠的聚合节点,并增加可信模型在聚合中的贡献占比,保证数据共享结果的可靠性.

④ 在公开数据集上进行对比实验,验证了 STDPPS-SL 方法的可用性及安全性.

1 背景知识

1.1 群体学习

群体学习(swarm learning,SL)[15]是研发机构 Hewlett Packard Labs 开发的继联邦学习后的新一代隐私保护分布式机器学习方法. 与联邦学习方法相比,群体学习方法在网络中取消了中心服务器,转而通过设计模型聚合算法并引入区块链,实现去中心化且参与方权益平等的协同模型训练. 本文使用群体学习作为实现时空数据共享的网络架构和方法基础. 群体学习示意图如图 1 所示.

1.2 差分隐私

差分隐私(differential privacy,DP)[16]是一种基于随机噪声扰动的隐私保护方法,其通过对查询结果附加随机噪声,减少输入数据集中任意数据对最终查询结果的影响程度. 经过差分隐私处理的查询结果仅包含输入数据集中数据普遍具备的统计学性质,数据自身特征则将被严格控制. 本文使用差分隐私作为实现时空数据共享过程中参与方模型隐私保护、对抗针对模型参数的分析方法的技术手段.

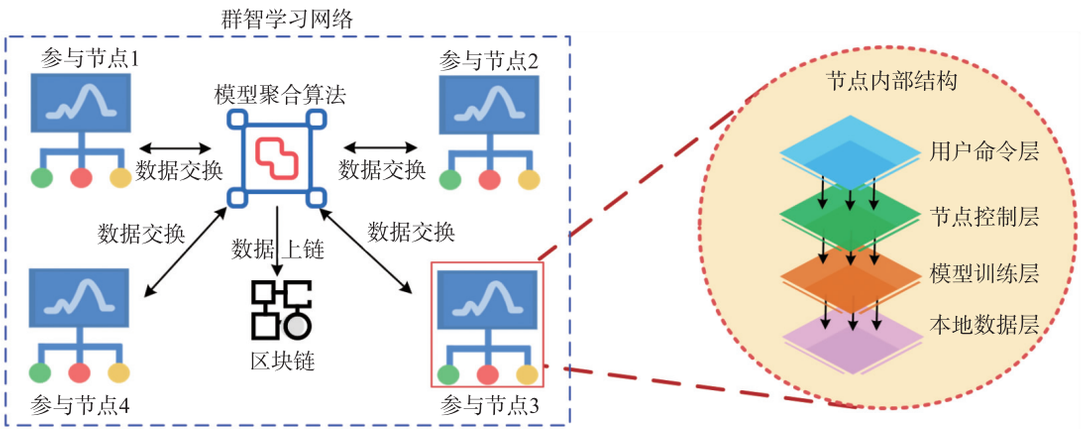


图 1　群体学习示意图

Fig. 1　Schematic diagram of swarm learning

定义 1　严格差分隐私[17]

给定两相邻数据集 D、D',对于一随机算法 M 及其值域 $R(M)$,令 $O\in R(M)$,若满足下式,则称算法 M 满足严格差分隐私.

$$\frac{\Pr[M(D)\in O]}{\Pr[M(D')\in O]}\leqslant e^{\varepsilon}, \tag{1}$$

式中,概率 Pr 为事件发生的概率,由随机算法 M 控制;隐私预算 ε 为反映隐私保护强度的参数指标,ε 越小,噪声扰动越强,隐私保护强度越高.

定义 2　全局敏感度[18-19]

对于一确定实值函数 f,D 和 D'为相邻数据集,则函数 f 的全局敏感度 Δf 为:

$$\Delta f=\max_{D,D'}\|f(D)-f(D')\|_1. \tag{2}$$

1.3　可认证随机函数[20]

可认证随机函数(verifiable random function,VRF)是一种特殊的随机数生成器,给定随机数种子 seed,公私钥密钥对(p_k,s_k),其能够生成一个随机数 random 及其证明 proof. 使用公钥 p_k 可以验证 random 是否是由 p_k 对应的 s_k 生成的. 对于一个随机数 random,其由且仅由私钥 s_k 及种子 seed 决定. 可认证随机函数是本文所提 STDPPS-SL 方法实现去中心化数据共享的关键技术.

2　时空数据隐私保护共享的群体学习(STDPPS-SL)方法设计

为了在推动时空数据共享、优化共享效果的同时兼顾参与方经济效益及合法权益,本文提出时空数据隐私保护共享的群体学习(STDPPS-SL)方法. STDPPS-SL 方法示意图如图 2 所示,阶段示意图如图 3 所示.

由图 2 可知,STDPPS-SL 方法主要由群体学习网络及区块链两部分组成. 群体学习网络是由参与方以节点形式组成的具有去中心化特性的分布式数据共享网络,在方法中负责实现无需直接传输且参与方权益平等的时空数据共享;区块链是一种分布式数据库,其存储的数据具备防篡改及可追溯特性,在方法中负责记录重要的过程信息,保证数据共享结果可信.

由图 3 可知,STDPPS-SL 方法主要包括 4 个阶段. 阶段 1:参与方本地模型训练阶段,实现参与方使用私有时空数据训练满足差分隐私的本地模型,主要提出 t-dpSGD 算法;阶段 2:网络全局模型生成阶段,实现基于各参与方本地模型参数生成性能可靠的网络全局模型参数,以及领导该过程的聚合节点的选择,主要提出自适应模型聚合算法和聚合节点选择算法;阶段 3:参与方可信度调整阶段,实现基于各参与方对数据共享贡献调整其可信度分数,主要提出参与方打分机制;阶段 4:参与方本地模型更新阶段,实现各参与方使用生成的网络共享模型代替原本的本地模型,用于新一轮的 STDPPS-SL 过程.

2.1　参与方本地模型训练阶段

为了在时空数据共享过程中,完成参与方本地模型训练的同时保护参与方私有数据所有权,避免隐私

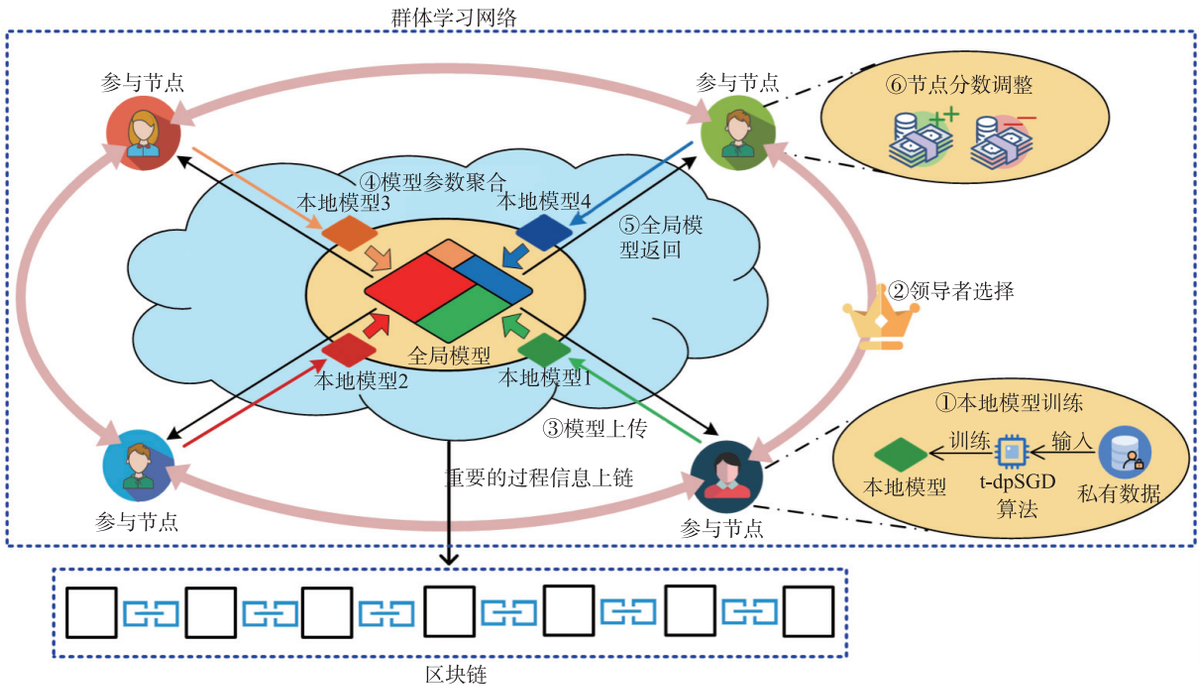


图 2　STDPPS-SL 方法示意图

Fig. 2　Schematic diagram of STDPPS-SL

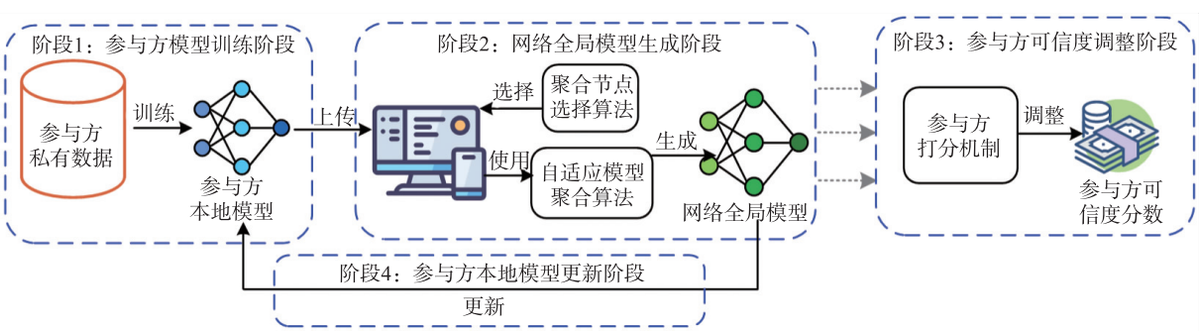


图 3　STDPPS-SL 阶段示意图

Fig. 3　Schematic diagram of STDPPS-SL phases

泄露,提出基于 t 分布的差分隐私随机梯度下降(t-dpSGD)算法.

2.1.1　t-dpSGD 算法

t-dpSGD 算法的核心思想:以传统的随机梯度下降模型训练方法为基础,通过梯度裁剪控制梯度变化量上界,使用符合 t 分布的随机噪声实现模型参数扰动,从而训练符合严格差分隐私的机器学习模型. t-dpSGD 算法如算法 1 所示.

算法 1:t-dpSGD 算法

输入:节点本地模型参数 θ,本地数据集 LocalData,梯度裁剪上界 C,损失函数 $L(\theta)$,训练子集数据量 n,学习率 η

输出:隐私保护模型参数 θ'

1. 将数据集 LocalData 划分为多个含有 n 条数据的不相交子集,记为 $D_1,\cdots,D_j$
2. 随机选择 $D_i=\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$
3. FOR(x_i,y_i) IN D_i DO
4. $\quad G(x_i)=\nabla_\theta L(\theta,x_i,y_i)$
5. $\quad g(x_i)=\dfrac{G(x_i)}{\max\left(1,\dfrac{\|G(x_i)\|_2}{C}\right)}$　//梯度裁剪
6. END FOR

7. $\bar{g}=\frac{1}{n}\sum_{i=1}^{n}(g(x_i))$

8. $\tilde{g}=\bar{g}+N$,噪声 N 概率密度函数 $f(x)=\frac{\Gamma(1)}{\Gamma\left(\frac{1}{2}\right)\sqrt{\pi}}(1+x^2)^{-1}$

9. $\theta'=\theta-\eta\tilde{g}$

10. RETURN θ'

2.1.2 t-dpSGD 算法的全局敏感度上界与差分隐私特性证明

定理 1 令 $h(x)$ 表示算法 1 步骤 4~7 的梯度裁剪过程,则 $h(x)$ 的全局敏感度 $\Delta h=2C$.

证明 设参数 $\boldsymbol{\theta}$ 为 n 维向量,即 $\boldsymbol{\theta}=(\theta_1,\theta_2,\cdots,\theta_n)$,则梯度向量 $\boldsymbol{G}=(G_1,G_2,\cdots,G_n)$.

由算法 1,有:

$$\bar{g}=G\cdot\frac{k}{\|G\|_2}=k\cdot\frac{g}{\|G\|_2}. \tag{3}$$

取梯度向量 $\boldsymbol{g}$ 中任意元素 g_i,可知:

$$\bar{g}_i=\frac{G_i}{\|G\|_2}=\frac{G_i}{\sqrt{G_1^2+G_2^2+\cdots+G_i^2+\cdots+G_n^2}}\leqslant\frac{G_i}{\sqrt{G_i^2}}=1, \tag{4}$$

故有 $\bar{g}_i\leqslant C$.

取两邻接数据子集 D 与 D' 分别执行算法 1 步骤 4~7,可得两梯度:

$$\bar{g}^1=(\bar{g}_1,\bar{g}_2,\cdots,\bar{g}_n),\bar{g}^2=(\bar{g}_1,\bar{g}_2,\cdots,\bar{g}'_n), \tag{5}$$

则有:

$$\max\|h(D)-h(D')\|_1=(0,0,\cdots,\tilde{g}_n-\tilde{g}'_n)=(0,0,\cdots,\frac{1}{n}\sum\bar{g}_n-\frac{1}{n}\sum\bar{g}'_n)\leqslant 2C. \tag{6}$$

定理 1 得证.

定理 2 设定理 1 中 $h(x)$ 的值域 $R\in[a,b]$,则算法 1 符合 $\left(\ln\left(1+\frac{4aC+4C^2}{1+a^2}\right),0\right)$ 差分隐私.

证明 取相邻数据集 D 与 D',使用 r 与 r' 分别表示两数据集内的记录. 假设两数据集存在唯一的相异记录,将其分别表示为 r_n 与 r'_n. 记算法 1 过程为 M,有:

$$\frac{\Pr[M(D)=O]}{\Pr[M(D')=O]}=\prod_{i=1}^{n}\frac{\Pr[M(r_i)=O_i]}{\Pr[M(r'_i)=O_i]}=\frac{\Pr[M(r_i)=O]}{\Pr[M(r'_i)=O]}. \tag{7}$$

因:

$$M(x)=h(x)+N, \tag{8}$$

设:

$$h(d_n)=O-S,h(d'_n)=O-S-|d_n-d'_n|, \tag{9}$$

联立式(7)、(8)和(9),有:

$$\frac{\Pr[M(d_n)=O]}{\Pr[M(d'_n)=O]}=\frac{\Pr[N=S]}{\Pr[N=S+|d_n-d'_n|]}=\frac{\int_S^{S+\Delta x}f(x)\,\mathrm{d}x}{\int_{S+\Delta d}^{S+\Delta d+\Delta x}f(x)\,\mathrm{d}x}=\frac{f(N)}{f(N+\Delta d)}=\frac{1+(N+\Delta d)^2}{1+N^2}\leqslant$$

$$\frac{1+N^2+4NC+4C^2}{1+N^2}=1+\frac{4NC+4C^2}{1+N^2}.$$

此时有:

$$\frac{\Pr[M(D)=O]}{\Pr[M(D')=O]}\leqslant e^{\ln\left(1+\frac{4aC+4C^2}{1+a^2}\right)}. \tag{10}$$

定理 2 得证.

2.2 参与方可信度分数调整阶段与参与方模型聚合阶段

群体学习与联邦学习最大的不同在于群体学习具有去中心化特性,取消了网络中的参数服务器. 为

了保证去中心化网络中的参与方诚实可信、协同实现时空数据共享过程,设计参与方打分机制以量化参与方可信程度;同时基于该分数设计聚合节点选择算法和自适应模型聚合算法,保证最终共享模型稳定可靠,性能优秀.

2.2.1 参与方打分机制

参与方打分机制的核心思想:基于参与方本地时空数据综合质量为参与方进行一个初始分数赋值,并根据参与方在后续为数据共享做出的贡献对分数进行调整. 参与方打分机制由两部分构成,分别为分数初始化算法和分数累计算法.

(1)分数初始化算法

分数初始化算法对参与方本地私有数据的综合价值进行评估:

$$\text{score}=\sum_{t=1}^{n}\alpha\cdot v_t\cdot i_t, \tag{11}$$

式(11)中,α 为权重系数,用于调整最终输出的分数量级;v_t 与 i_t 分别对应类别为 t 的时空数据的参与方本地数据量及预设的该类数据重要程度. 分数初始化算法认为参与方本地数据量越大、数据重要程度越高,其越可信. 事实上,提供越多的本地时空数据,提供的时空数据越重要,参与方对时空数据共享的贡献就越大,这与时空数据共享的目的相符.

(2)分数累计算法

分数累计算法的核心思想:根据参与方在时空数据共享过程中的工作量动态调整参与方分数. 普通参与方对共享模型的贡献集中在提供的本地模型参数,故工作量可由本地模型质量近似表示. 被选作聚合节点的参与方对共享模型的贡献则集中于执行自适应模型聚合算法. 分数累计算法如算法 2 所示.

算法 2:分数累计算法

输入:参与方列表 PList,合格值 R_1、R_2,公共测试集 D_{test},测试数量 n,分数调整粒度 v_1、v_2,阶梯值 t
输出:参与方分数 score

```
1. FOR 参与方 IN PList DO
2. IF 参与方不是聚合节点 THEN
3.     从 D_test 中随机取 n 条数据,使用交叉熵计算参与方本地模型质量 q
4.     IF q≤R_1 THEN
5.         score+=v_1
6.     ELSE
7.         score-=v_1 *(该参与方本地模型不及格轮次%t)
8.     ENF IF
9. ELSE
10.    IF 接受参与方生成共享模型的节点数量 number≥R_2 THEN
11.        score+=v_1+v_2 *(number/len(PList))
12.    ELSE
13.        score-=v_2 *(1-(number/len(PList)))+v_1 *(该参与方生成共享模型不及格轮次%t)
14.    END IF
15. END WHILE
16. RETURN score
```

通过分数累计算法调控参与方信任度分数,可以有效量化参与方的可信程度,为聚合节点选择与自适应模型聚合提供基础,同时也可以有效激励参与方进一步共享本地私有数据,提供性能更为优异的本地模型或共享模型.

在实际应用过程中,聚合节点不会将质量低于合格值的模型聚合入共享模型,同时当拒绝模型的参与方数量达到预设阈值时,STDPPS-SL 网络将会放弃本轮全局模型,并对领导者做出较大的惩罚. 可将参与方分数及其变化值以及算法 2 过程信息,通过打包上链记录在区块链中,从而保证分数不可篡改及可追溯.

2.2.2 聚合节点选择算法

聚合节点是 STDPPS-SL 方法中模型参数聚合阶段动态选择的参与方节点,其负责收集各参与方本地模型参数,执行自适应模型聚合算法生成共享模型,并将共享模型返回给各参与方以实现模型更新.

聚合节点选择算法的核心思想:可信程度越高的参与方,越能更好地执行聚合节点责任;但同时仅以参与方可信程度为衡量标准,会导致聚合节点选择范围局限,部分参与方垄断聚合节点身份,故结合参与方分数与随机值作为选择权重,权重最大者即为聚合节点. 聚合节点选择算法如算法 3 所示.

算法 3:聚合节点选择算法

输入:参与方列表 PList,种子 seed
输出:聚合节点 AggNode
1. FOR 参与方 IN PList DO
2. random、proof = $\mathrm{VRF}_{\text{参与方私钥sk}}$(seed)
3. weight = random+参与方当前可信分数
4. 参与方将 random、proof、weight 等相关信息打包上链
5. END FOR
6. 选择 weight 最大的参与方作为聚合节点 AggNode
7. RETURN AggNode

在进行领导者选择前,各参与方需要拥有自己的密钥对(p_k, s_k),此需求可以通过区块链网络实现. 算法 3 中出现的种子 seed,采用如下的方式产生:

$$\mathrm{VRF}_{sk_{r-1}}(\mathrm{seed}_{r-1} \parallel r \parallel T) \rightarrow (\mathrm{seed}_r, \mathrm{proof}), \tag{12}$$

式中,r 代表当前 STDPPS-SL 方法训练轮数,sk_{r-1}为 $r-1$ 轮聚合节点的私钥,T 为当前时间. 进行首轮执行时,seed 置为预先设定的填充字节连接当前时间所得的字符串.

2.2.3 自适应模型聚合算法

自适应模型聚合算法的核心思想:参与方可信程度越高,其本地模型越可靠,故提高可信模型被共享模型学习的程度、降低不可信模型被共享模型学习的程度,可以有效提高作为数据共享结果的共享模型自身的可靠性.

STDPPS-SL 方法依照下式进行参与方本地模型聚合以生成共享模型:

$$\theta_{\text{share}} = \sum_{i=1}^{\text{len(PList)}} \frac{\text{score}_i}{\text{score}_{\text{sum}}} \theta_i. \tag{13}$$

由式(13)可知,共享模型由各参与方本地模型按照不同的权重线性组合而成. 与将各参与方本地模型参数均值作为整体模型的 FedAVG 等传统模型聚合方法相比,自适应模型聚合算法会依照参与方分数自适应地调整其模型被整体模型学习的比例,对于分数较高(更可信)的参与方,其模型会被以较大的比例整合进整体模型中.

3 实验分析及讨论

3.1 实验设置

实验环境:实验环境硬件配置为四核处理器 11th Gen Intel(R) Core(TM) i5-11300H、GTX 1080T GPU、16 GB RAM. 在 Ubuntu 20.04.3 系统下,使用 PyTorch、TensorFlow 等包实现本文设计方法.

实验数据:实验使用北京市 2008—2022 年历史天气数据作为时空数据集. 记录由日期、天气状况、最低/最高气温及风力风向 4 类特征构成,取 12 800 条数据为训练集,预测任务为基于目标日期前两天的天气状况预测目标日期当天天气状况.

3.2 实验过程与实验分析

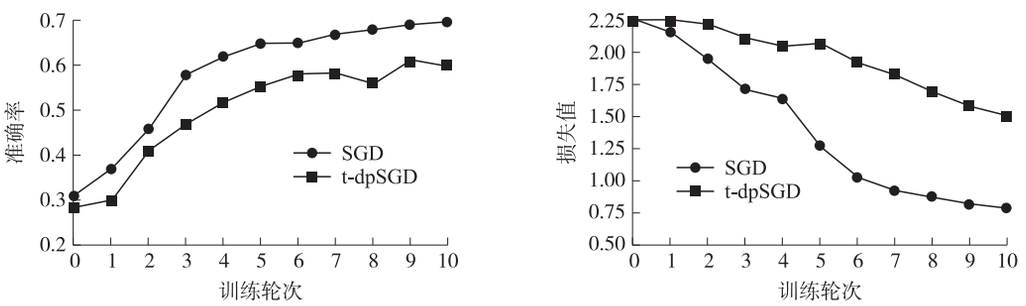
为了从隐私保护性、方法效率、模型训练性能等多个方面分析本文所提方法的综合性能,设计 3 个对比实验:(1)对比本文所提 t-dpSGD 算法与原始随机梯度下降算法在准确率和损失值指标上的表现,同时对比在不同隐私预算上限下 t-dpSGD 算法的实际隐私预算,证明 t-dpSGD 算法能够在满足模型训练需求的同时实现隐私保护;(2)对比本文所提聚合节点选择算法与随机选择聚合节点方法的运行所需时间,证

明本文所提聚合节点选择算法能够在保证数据共享结果可靠性的前提下,满足多参与方时空数据共享的效率要求;(3)对比本文所提 STDPPS-SL 方法与联邦学习、集中式学习方法在准确率指标上的表现,证明本文所提方法能够兼顾安全性与可用性,实现参与方隐私保护的时空数据共享.

3.2.1　t-dpSGD 算法对比实验

实验 1 使用长短时记忆网络(long short-term memory,LSTM)模型,首先对比原始随机梯度下降(SGD)算法与 t-dpSGD 算法在模型训练方面的性能差异,探究 t 分布随机噪声扰动对模型性能的影响;其次,对比在不同隐私预算上界限制下使用 t-dpSGD 算法训练的模型参数实际隐私预算的增长情况,以分析该方法实现差分隐私的实际效果.

首先,分别使用 SGD 算法和 t-dpSGD 算法进行模型训练,并对比模型准确率与损失值. 实验结果如图 4 所示.



(a) 准确率对比　　(b) 损失值对比

图 4　t-dpSGD 算法对比实验图

Fig. 4　Comparative experimental graph of t-dpSGD algorithm

其次,在不同隐私预算上限下使用 t-dpSGD 算法进行模型训练,记录模型参数的实际隐私预算. 实验结果如图 5 所示.

实验结果分析:

(1)t-dpSGD 算法能够兼顾神经网络模型训练与模型参数差分隐私. 由图 4 可知,t-dpSGD 与 SGD 准确率相差约 10%,平均损失相差约 0.8,这是因为 t-dpSGD 为了实现参数差分隐私引入的随机噪声影响了模型准确率,但其能够实现模型性能与差分隐私的平衡.

(2)在 t-dpSGD 算法运行过程中,实际隐私预算与训练轮数呈近似正比关系. 由图 5 可知,随着训练轮数的增加,实际隐私预算也同步上升.

3.2.2　聚合节点选择算法对比实验

实验 2 使用基于椭圆曲线的可认证随机函数实现聚合节点选择算法,与随机选择聚合节点对比在不同参与节点数量下的运行所需时间. 实验结果如图 6 所示.

实验结果分析:聚合节点选择算法在 500 个参与节点的环境下运行时间约为 530 s,能够满足多参与方时空数据共享的效率要求.

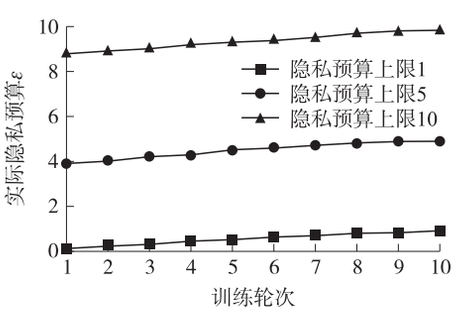


图 5　t-dpSGD 实际隐私预算图

Fig. 5　t-dpSGD actual privacy budget graph

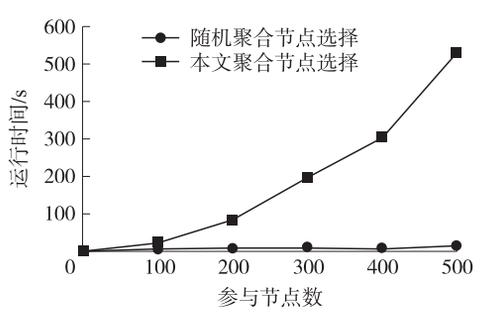


图 6　聚合节点选择算法对比实验图

Fig. 6　Comparative experimental graph of aggregation node selection algorithm

3.2.3 STDPPS-SL 方法对比实验

实验 3 使用 LSTM 模型,探究 STDPPS-SL、联邦学习和集中式学习在时空数据隐私保护共享方面的性能差异. 实验结果如图 7 所示.

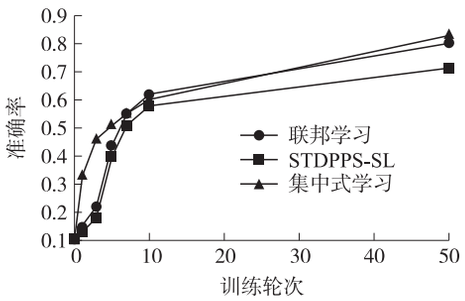


图 7 STDPPS-SL 方法对比实验图

Fig. 7 Comparative experimental graph of STDPPS-SL

实验结果分析:STDPPS-SL 方法的模型训练性能接近联邦学习方法. 由图 7 可知,二者模型准确率差值最大约为 9%,同时,由于集中式学习相当于各参与方上传本地私有时空数据,在能够直接获取到数据的前提下,其表现优于联邦学习与 STDPPS-SL. STDPPS-SL 方法由于为了实现模型参数差分隐私对参数进行了噪声扰动,这使得其模型表现略低于联邦学习与集中式学习方法,其付出了约为 7% 的准确率作为代价,实现了模型参数隐私保护,同时可以防御重构攻击等攻击方式,在保证模型可用性的前提下,兼顾了方法的安全性与隐私保护性.

4 结论

为了在推动时空数据共享、优化共享效果的同时兼顾参与方经济效益及合法权益,本文提出了一种时空数据隐私保护共享的群体学习(STDPPS-SL)方法,该方法基于群体学习构建多参与方时空数据共享网络,提出保护参与方本地私有数据敏感信息的 t-差分隐私随机梯度下降算法,设计参与方打分系统量化其可信程度,并结合参与方分数设计聚合节点选择算法与自适应模型聚合算法,提高数据共享结果的性能与可靠性. 在公开数据集上的实验结果表明,STDPPS-SL 方法具备安全性与可用性. 今后的研究将考虑:(1)设计基于群体学习与区块链的时空数据市场化共享及协同分析体系;(2)设计一种结合边缘计算、云计算与隐私计算的时空数据共享方法.

[参考文献]

[1] 杨雅萍,姜侯,孙九林. 科学数据共享实践:以国家地球系统科学数据中心为例[J]. 地球信息科学学报,2020,22(6):1358-1369.

[2] FAN L,WANG L. Secure sharing of spatio-temporal data through name-based access control[C]//Proceedings of the IEEE INFOCOM 2021—IEEE Conference on Computer Communications Workshops(INFOCOM WKSHPS). Vancouver,Canada:IEEE,2021:1-7.

[3] ZHAO T T,LIU W Z,DI X L,et al. Research on the way of sharing geographic information data in disaster management[J]. The international archives of the photogrammetry,remote sensing and spatial information sciences,2022,48(3):103-109.

[4] MCMAHAN B,MOORE E,RAMAGE D,et al. Communication-efficient learning of deep networks from decentralized data[C]//Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. Ft. Lauderdale,USA:PMLR,2017:1273-1282.

[5] YANG Q,LIU Y,CHEN T,et al. Federated machine learning:concept and applications[J]. ACM transactions on intelligent systems and technology,2019,10(2):1-19.

[6] KAIROUZ P,MCMAHAN B,AVENT B,et al. Advances and open problems in federated learning[J]. Foundations and trends in machine learning,2021,14(1):1-121.

[7] GRASER A,HEISTRACHER C,PRUCKOVSKAJA V. On the role of spatial data science for federated learning[C]//Proceedings of the Spatial Data Science Symposium 2022. Virtual. USA:eScholarship Publishing,2022:1-8.

[8] ASHISH K S. Advancing location privacy in urban networks:a hybrid approach leveraging federated learning and geospatial semantics[J]. International journal of information and cybersecurity,2023,7(1):58-72.

[9] CHUNG P,TAEKYOON C,TAESAN K,et al. FedGeo:Privacy-preserving user next location prediction with federated learning[C]//Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems. Hamburg,Germany:ACM,2023:13-16.

[10] 蒋伟进,韩裕清,吴玉庭,等. 基于边缘计算的环境监测自适应联邦学习算法[J]. 电子学报,2023,51(11):3061-3069.
[11] JONAS G,HARTMUT B,HANNAH D,et al. Inverting gradients-how easy is it to break privacy in federated learning? [C]// Proceedings of the 34th International Conference on Neural Information Processing Systems. Virtual. USA:Curran Associates Inc,2020:16937-16947.
[12] ZHANG Y H,JIA R X,PEI H Z,et al. The secret revealer:generative model-inversion attacks against deep neural networks [C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Seattle,USA:IEEE,2020:250-258.
[13] JOSHUA C Z,ATUL S,Elkordy A,et al. LOKI:large-scale data reconstruction attack against federated learning through model manipulation[C]//Proceedings of the 2024 IEEE Symposium on Security and Privacy(SP). San Francisco,USA:IEEE,2023:1287-1305.
[14] KANG H Y,JI Y R,ZHANG S X. Enhanced privacy preserving for social networks relational data based on personalized differential privacy[J]. Chinese journal of electronics,2022,31(4):741-751.
[15] WARNAT-HERRESTHAL S,SCHULTZE H,SHASTRY K L,et al. Swarm learning for decentralized and confidential clinical machine learning[J]. Nature,2021,594(7862):265-270.
[16] DWORK C,ROTH A. The algorithmic foundations of differential privacy[J]. Foundations and trends in theoretical computer science,2013,9(3):211-407.
[17] JOSEPH F,WANG W,CHEN H N,et al. Differential privacy in health research:a scoping review[J]. Journal of the American medical informatics association,2021,28(10):2269-2276.
[18] WANG Y L,WANG Q,ZHAO L C,et al. Differential privacy in deep learning:privacy and beyond[J]. Future generation computer systems,2023,148:408-424.
[19] 康海燕,王骁识. 基于数据特征相关性和自适应差分隐私的深度学习方法研究[J]. 电子学报,2024,52(6):1963-1976.
[20] MICALI S,RABIN M,VADHAN S. Verifiable random functions[C]//Proceedings of the 40th Annual Symposium on Foundations of Computer Science. New York,USA:IEEE,1999:120-130.

[责任编辑:丁 蓉]