

基于边缘计算的环境监测自适应联邦学习算法

蒋伟进^{1,2,3}, 韩裕清^{1,3}, 吴玉庭^{1,3}, 周 为^{1,3}, 陈艺琳^{1,3}, 王海娟^{3,4}

(1. 湖南工商大学计算机学院, 湖南长沙 410205; 2. 武汉理工大学计算机与人工智能学院, 湖北武汉 430070;
3. 湘江实验室, 湖南长沙 410205; 4. 湖南工商大学前沿交叉学院, 湖南长沙 410205)

摘要: 针对环境监测领域边缘设备资源不平衡、通信延迟以及模型质量不高的问题, 本文提出一种基于边缘计算的环境监测自适应联邦学习算法. 该算法旨在利用边缘设备进行数据处理, 并根据各个设备的资源限制调整全局模型的聚合频率, 以更好地适应不同的监测环境. 通过考虑边缘设备之间的资源差异, 算法采用了动态优化迭代频率的策略, 以提升模型的训练效果. 与传统的固定迭代频率相比, 该算法的调整策略更加灵活, 能够更好地适应不同的数据分布和参与方特征. 通过大量实验评估, 并利用与同类算法 CNN-FL (Convolutional Neural Networks-Federated Learning), FedAvg (Federated Averaging) 和 HFEL (Hierarchical Federated Edge Learning) 的比较, 本文提出的算法在算法性能和经济成本方面具有显著优势, 这种算法为环境监测提供了一种高效、安全和可扩展的数据分析和模型建立方法, 有助于推动环境监测能力的提升.

关键词: 环境监测; 自适应联邦学习; 边缘计算; 模型聚合; 优化算法

基金项目: 国家自然科学基金 (No.61772196, No.72088101); 湖南省自然科学基金 (No.2020JJ4249)

中图分类号: TP301

文献标识码: A

文章编号: 0372-2112(2023)11-3061-09

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20230504

Federated Learning Scheme for Environmental Monitoring Based on Edge Computing

JIANG Wei-jin^{1,2,3}, HAN Yu-qing^{1,3}, WU Yu-ting^{1,3}, ZHOU Wei^{1,3}, CHEN Yi-lin^{1,3}, WANG Hai-juan^{3,4}

(1. School of Computer Science, Hunan University of Technology and Business, Changsha, Hunan 410205, China;

2. School of Computer Science and Artificial Intelligence, Wuhan University of Technology, Wuhan, Hubei 430070, China;

3. Xiangjiang Laboratory, Changsha, Hunan 410205, China;

4. College of Frontier Intersection, Hunan University of Technology and Business, Changsha, Hunan 410205, China)

Abstract: Aiming at the problems of unbalanced edge device resources, communication delay and low model quality in the field of environmental monitoring, this paper proposes an adaptive federated learning algorithm for environmental monitoring based on edge computing. This algorithm aims to use edge devices for data processing, and according to each the resource limitation of the device adjusts the aggregation frequency of the global model to better adapt to different monitoring environments. By considering the resource differences between edge devices, the algorithm adopts a strategy of dynamically optimizing the iteration frequency to improve the training effect of the model. Compared with the traditional fixed iteration frequency, the adjustment strategy of this algorithm is more flexible and can better adapt to different data distribution and participant characteristics. Through a large number of experimental evaluations, and using the same algorithm convolutional neural networks-federated learning (CNN-FL), federated averaging (FedAvg) and hierarchical federated edge learning (HFEL), the algorithm proposed in this paper has significant advantages in algorithm performance and economic cost. This algorithm provides an efficient, safe and reliable method for environmental monitoring. Expanded approach to data analysis and modeling to help drive improvements in environmental monitoring capabilities.

Key words: environmental monitoring; adaptive federated learning; edge computing; model polymerization; optimization algorithm

Foundation Item(s): National Natural Science Foundation of China (No.61772196, No.72088101); Hunan Provincial Natural Science Foundation (No.2020JJ4249)

1 引言

在5G网络迅速发展下,许多智能设备快速融入并应用环境监测领域,例如:智能空气质量监测器^[1]、无人机^[2]、智能水质监测器^[3]、智能农业等,同时,这些智能设备也有了强大计算资源、人机交互能力以及低延时等需求,然而,将所有的环境监测数据传输到中央服务器进行集中式处理并不是理想的选择.因此,作为人工智能(Artificial Intelligence, AI)核心技术^[4,5]的联邦学习(Federated Learning, FL)^[6]成为了减少大规模数据处理的有效手段.

在环境监测领域,为了高效应用联邦学习,提高模型准确性和收敛速度,以更好地适用于环境监测任务,研究者们提出了多种联邦学习优化方法.文献[7]提出了一种基于无人机(Unmanned Aerial Vehicle, UAV)的感知框架,将UAV和地面节点组合起来,利用联邦学习和无人机技术来监测和评估城市中的空气质量,文献[8]提出了一种基于双向门控循环单元(Bidirectional Gated Recurrent Unit, BGRU)模型的联邦学习方法,用于预测智能城市中的空气质量,文献[9]提出一种基于联邦学习的农作物产量预测方法,通过神经网络模型提高预测准确性.

然而,现有研究虽然在特定领域存在一定优势,但在联邦学习优化算法中并未综合考虑资源约束、移动设备异构性、数据分布的非独立同分布等问题.此外,环境监测涉及大量的传感器和设备,这些设备产生的数据量巨大.如何利用边缘计算系统有效地收集、存储和处理这些数据仍是一个挑战.与此同时,环境监测系统需要适应不断变化的环境条件和监测需求,快速集成新设备和传感器来适应不同的监测任务和环境条件.因此本文将综合考虑这些问题,提出新的环境监测自适应联邦学习算法,更好地适应实际场景中的环境监测需求,并在其他多个领域具有潜在适用性.

2 基于边缘计算的环境监测自适应联邦学习算法

2.1 环境监测自适应联邦学习算法

在环境监测领域,处理大量敏感数据常常面临一些困难.这是因为数据无法轻易传输到中心服务器进行处理,同时网络带宽可能存在限制.此外,不同地区或不同传感器所收集到的数据可能存在差异,这种差异性可能对模型的训练和性能产生负面影响.因此,本文设计了一种边缘端主导的环境监测自适应联邦学习算法,边缘设备能根据实际环境监测需求和数据特征,动态的调整学习参数,这种动态调整能力使得监测设备可以根据环境条件的变化和监测任务的要求,优化模型的性能和效果.例如,农业监测系统可以充分利用

分布在农田中的边缘设备,实现本地化的模型训练和数据处理,使得模型在不同农业环境下具有更好的适应性和泛化能力.因此农业监测环境可以通过边缘计算系统中自适应联邦学习的协助,实现在边缘计算环境下的植物病害识别,为农作物健康管理和病害防治提供有效支持.

2.2 损失函数

假设环境监测有 N 个边缘节点,本地数据集分别被视为 $D_1, D_2, \dots, D_i, \dots, D_N$,以边缘节点 i 的数据集 D_i 为例,它的损失函数为

$$F_i(\mathbf{w}) = \frac{1}{|D_i|} \sum_{j \in D_i} f_j(\mathbf{w}) \quad (1)$$

对于 $i \neq i', D_i \cap D_{i'} = \emptyset$,边缘节点的全局损失函数为

$$F(\mathbf{w}) = \frac{1}{|\bigcup_i D_i|} \sum_{j \in \bigcup_i D_i} f_j(\mathbf{w}) = \frac{1}{D} \sum_{i=1}^N D_i F_i(\mathbf{w}) \quad (2)$$

$D = \sum_{i=1}^N D_i$ 进行模型训练获得最小化损失函数,由于机器学习模型复杂,为了找到最小化损失函数的参数,通常采取梯度下降法找到最小化 $F(\mathbf{w})$.

$$\mathbf{w}^* = \operatorname{argmin} F(\mathbf{w}) \quad (3)$$

2.3 分布式梯度下降

分布式梯度下降算法的模型更新分为两个部分,包含参与方的本地更新步骤与中心方的全局聚合步骤,参与方的每个 i 节点都存在着一个局部模型参数 $\mathbf{w}_i(t), \mathbf{w}_i(t)$ 中的 $t=0, 1, \dots, n$ 表示着索引迭代, $t=0$ 时,所有 i 节点在本地初始化为同样的值,当 $t \neq 0$ 时,根据局部损失函数的梯度下降更新规则在前一次迭代的基础上计算新的 $\mathbf{w}_i(t)$ 值,其中,节点在做局部损失函数梯度下降的步骤称为本地更新^[10].中心方根据各个参与方发来的本地更新参数做加权平均,为参与方的本地更新做最新的加权平均值.

在前一次迭代完成全局聚合后,对参数执行本地更新操作,对于每个边缘 i 节点,具体更新规则如下:

$$\mathbf{w}_i(t) = \tilde{\mathbf{w}}_i(t-1) - \eta \nabla F_i(\tilde{\mathbf{w}}_i(t-1)) \quad (4)$$

在进行全局聚合时,

$$\mathbf{w}(t) = \frac{1}{D} \sum_{i=1}^N D_i \mathbf{w}_i(t) \quad (5)$$

全局聚合期间,每个 i 节点执行 τ 步本地更新,每处节点执行 T 次本地迭代,定义 T 为 τ 的整数倍,其中, τ 为两个全局聚合之间的本地更新迭代次数, T 为每个节点的本地更新步骤总数.

2.4 问题模型化

每当数据分布在大量边缘节点上时,联邦学习在处理这些数据会消耗不小的时间资源与能量^[11],因此,

需要在系统正常运行的基础上限制时间资源,首先本文将在基于梯度下降的分布式学习方法基础上缩小问题,确定 T 和 τ 的最优值,从而在给定时间资源预算下最小化全局损失函数,由于 T 是 τ 的整数倍,全局聚合总数 $K = T/\tau$. 文章定义:

$$\mathbf{w}^f = \underset{\mathbf{w} \in \{\mathbf{w}(k\tau); k=0,1,\dots,K\}}{\operatorname{argmin}} F(\mathbf{w}) \quad (6)$$

其中, \mathbf{w}^f 为学习结束后获得的最终模型参数,为了计算式(6)中全局损失函数 $F(\mathbf{w})$, 每个 i 节点首先计算 $F_i(\mathbf{w})$, 再将结果发送给聚合器进行进一步处理. 由于每个节点只知道第 k 次全局聚集后的 $\mathbf{w}(k\tau)$ 的值,因此在第 $(k+1)$ 次全局聚集时,节点 i 处的 $F_i(\mathbf{w}(K\tau))$ 将被发送回聚合器计算. 为了得到最后一个损失值 $F_i(\mathbf{w}(K\tau)) = F(\mathbf{w}(T))$, 在最后执行一轮局部和全局更新. 本文假设每个节点,无论是仅在最后一轮中计算局部损失,还是在其他所有轮中同时计算局部损失和梯度,所需的资源量都相同,因为损失和梯度计算通常基于相同的中间结果. 例如,用于计算神经网络梯度的反向传播方法需要一个正向传播过程,该过程实质上是作为中间步骤获得损耗^[12].

在模型训练时,局部模型训练消耗 c 个单位的时间资源,全局模型聚合时消耗 b 个单位时间资源,对于 T 和 τ ,消耗的时间资源总量为 $(T+1)c + (K+1)b$, 其中,为了便于计算 $F_i(\mathbf{w}(K\tau))$ 选择加 1.

记 P 为时间资源总预算,要求如下:

$$\min_{\tau, K \in \{1,2,3,\dots,K\}} F(\mathbf{w}^f), (T+1)c + (K+1)b \leq P, \quad T = K\tau \quad (7)$$

由于根据 K 与 τ 以及最终模型参数 \mathbf{w}^f 计算 $F(\mathbf{w}^f)$ 取决于许多因素,因此,我们将根据梯度下降的收敛性、全局聚合对收敛性的影响^[13], 以及现实中处于变化的 c 与 b 的资源消耗情况找到 $K, \tau, F(\mathbf{w}^f)$ 之间关系.

联邦学习的代价主要包括学习时间和能量消耗部分,通过计算模型迭代以及通信所需要的能量和时间可以估算出预计的经济成本. 具体如下:

(1) 时间开销: 本文在选取的 F 个客户端进行模型训练,客户端每次迭代所消耗的时间为 t_d , 全局通信过程中消耗的时间为 t_j , 每个客户端经过 τ 轮后学习时间 T_τ 为

$$T_\tau = \sum_{r=1}^{\tau} t_d^r + t_j \quad (8)$$

经过 T 轮后的 F 个客户端的总学习时间为

$$T_{\text{all}}(F, T) = \sum_{f=1}^F \sum_{r=1}^T T_\tau \quad (9)$$

(2) 能量开销: 客户端每次迭代所消耗的能量为 e_d , 全局通信消耗为 e_j , 经过 τ 轮后能量消耗 E_τ 为

$$E_\tau = \sum_{r=1}^{\tau} e_d^r + e_j \quad (10)$$

经过 F 轮后总能量消耗 E_{all} 为

$$E_{\text{all}}(F, \tau) = \sum_{f=1}^F \sum_{r=1}^T E_\tau \quad (11)$$

最后经济成本分析引入权重 $S \in [0, 1]$, 经济成本 M_{all} 为

$$M_{\text{all}} = T_{\text{all}}(F, T)S + (1-S)E_{\text{all}}(F, T) \quad (12)$$

2.5 收敛性分析

2.5.1 定义

本文将 T 次迭代分别分解为 K 个不同的间隔, 只有第一次与最后一次迭代拥有全局聚合, 并使用 $[k]$ 符号来表示迭代之间间隔 $[(k-1)\tau, k\tau]$, $k=1, 2, \dots, K$, 并且对于区间 $[k]$ 使用 $v_{[k]}(t)$ 表示一种辅助参数^[14], 参数满足中心化梯度下降:

$$v_{[k]}(t) = v_{[k]}(t-1) - \eta \nabla F(v_{[k]}(t-1)) \quad (13)$$

$v_{[k]}(t)$ 在区间 $[k]$ 开始处与 $\mathbf{w}(t)$ 同步, 即 $v_{[k]}((k-1)\tau) = \mathbf{w}((k-1)\tau)$, $\mathbf{w}(t)$ 为式(5)定义的局部参数平均值^[15], 因为在全局聚合或者是 $k=1$ 时的初始化是在迭代的 $(k-1)\tau$ 进行的, 因此对于所有的 i 节点有 $\tilde{\mathbf{w}}_i((k-1)\tau) = \mathbf{w}((k-1)\tau)$.

引理 1 $F(\mathbf{w})$ 是凸的, ρ -Lipschitz 以及 β -smooth 的.

证明可以从 $F(\mathbf{w})$ 的定义以及三角不等式来论证.

局部损失函数的梯度以及全局损失函数梯度之间存在一定差异, 这种差异与在不同节点数据的分布方式存在一定关联.

定义 1 梯度差异 对于任意 i 节点和 \mathbf{w} , 定义 δ_i 为 $\|\nabla F_i(\mathbf{w}) - \nabla F(\mathbf{w})\|$ 的上界, 即: $\|\nabla F_i(\mathbf{w}) - \nabla F(\mathbf{w})\| \leq \delta_i$, $\delta = \frac{1}{D} \sum_i D_i \delta_i$.

2.5.2 主要结果

下面定理给出了 $\mathbf{w}(t)$ 和 $v_{[k]}(t)$ 的差距上界.

定理 1 对于区间 $[k]$, $t \in [k]$, 拥有

$$\|\mathbf{w}(t) - v_{[k]}(t)\| \leq h(t - (k-1)\tau) \quad (14)$$

$$h(x) = \frac{\delta}{\beta} \left((\eta\beta + 1)^x - 1 \right) - \eta\delta x \quad (15)$$

而且, 因为 $F(\cdot)$ 是 ρ -Lipschitz 的, 那么有 $F(\mathbf{w}(t)) - F(v_{[k]}(t)) \leq \rho h(t - (k-1)\tau)$.

当 $t = (k-1)\tau$, 式(14)的上界为 0, 即进行全局聚合之后只进行一次本地更新时, 分布式梯度下降和中心梯度下降差别不大, 如果 $\tau = 1$, 那么 $t - (k-1)\tau$ 是 0 或 1, 那么式(14)的上界是确定值, 如果当 $\tau > 1$, $t - (k-1)\tau$ 可以取较大值. 当 $t - (k-1)\tau$ 很大时, 式(15)中的 $(\eta\beta + 1)^x$ 产生很大影

响, $\mathbf{w}(t)$ 和 $v_{[k]}(t)$ 的差距上界呈现指数增长, 另外, 局部梯度与全局梯度的差异越大, 间隙越大, 这种差异是由于在每次全局聚合之后的第二次局部更新开始时, 不同节点的局部梯度不同造成的. 在极端情况下, 当所有节点具有完全相同的数据样本时, 梯度会始终同并保持 $\delta=0$, 这种情况下, $\mathbf{w}(t)$ 和 $v_{[k]}(t)$ 会保持相等.

根据定理 1 得出下面引理.

引理 2 当下面条件均被满足时,

- (1) $\eta \leq \frac{1}{\beta}$
- (2) $\eta\varphi - \frac{\rho h(\tau)}{\tau\varepsilon^2} > 0$
- (3) $F(v_{[k]}(k\tau)) - F(\mathbf{w}^*) \geq \varepsilon$
- (4) $F(\mathbf{w}(T)) - F(\mathbf{w}^*) \geq \varepsilon$

对于 $\varepsilon > 0$, 论文定义 $\varphi = \omega\left(1 - \frac{\beta\eta}{2}\right)$, $\omega = \min_k \frac{1}{\|v_{[k]}((k-1)\tau) - \mathbf{w}^*\|^2}$, τ 次迭代后收敛上界为

$$F(\mathbf{w}(T)) - F(\mathbf{w}^*) \leq \frac{1}{T\left(\eta\varphi - \frac{\rho h(\tau)}{\tau\varepsilon^2}\right)} \quad (16)$$

定理 2 当 $\eta \leq \frac{1}{\beta}$, 有

$$F(\mathbf{w}^f) - F(\mathbf{w}^*) \leq \frac{1}{2\eta\varphi T} + \sqrt{\frac{1}{4\eta^2\varphi^2 T^2} + \frac{\rho h(\tau)}{\eta\varphi\tau}} + \rho h(\tau) \quad (17)$$

由式(17)可知, 引理 2 的条件 1 满足.

当 $\rho h(\tau) = 0$, 选取大于 0 条件下任意小的 ε , 这种情况下式(16)和式(17)的右边相等, 并且式(17)的结果能由引理 2 得出 $F(\mathbf{w}^f) - F(\mathbf{w}^*) \leq F(\mathbf{w}(T)) - F(\mathbf{w}^*)$. 在满足 $\rho h(\tau) > 0$ 时, 考虑式(17)的右边并使求解 ε_0 :

$$\varepsilon_0 = \frac{1}{T\left(\eta\varphi - \frac{\rho h(\tau)}{\tau\varepsilon_0^2}\right)} \quad (18)$$

得到:

$$\varepsilon_0 = \frac{1}{2\eta\varphi T} + \sqrt{\frac{1}{4\eta^2\varphi^2 T^2} + \frac{\rho h(\tau)}{\eta\varphi\tau}} \quad (19)$$

其中, 舍去负解, 因为 $\varepsilon > 0$, 式(18)中的分母大于 0, 在 $\varepsilon > \varepsilon_0$ 条件下满足引理 2 中的第二个条件.

假如存在 $\varepsilon > \varepsilon_0$ 满足引理 2 中的条件 3 和 4, 即满足引理 2 的全部条件. 将引理 2 代入到式(18)里, 有公式:

$$F(\mathbf{w}(T)) - F(\mathbf{w}^*) \leq \frac{1}{T\left(\eta\varphi - \frac{\rho h(\tau)}{\tau\varepsilon^2}\right)} < \frac{1}{T\left(\eta\varphi - \frac{\rho h(\tau)}{\tau\varepsilon_0^2}\right)} = \varepsilon_0$$

这就与引理 2 中第四个条件矛盾, 不存在 ε 同时满足条件 3 和 4, 即需要:

$$(1) \text{ 存在 } k, F(v_{[k]}(k\tau)) - F(\mathbf{w}^*) \leq \varepsilon_0$$

$$(2) F(\mathbf{w}(T)) - F(\mathbf{w}^*) \leq \varepsilon_0$$

即

$$m\left\{\min_{k=1,2,\dots,K} F(v_{[k]}(k\tau)); F(\mathbf{w}(T))\right\} - F(\mathbf{w}^*) \leq \varepsilon_0 \quad (20)$$

从定理 1 可知,

$$F(\mathbf{w}(k\tau)) \leq F(v_{[k]}(k\tau)) + \rho h(\tau), \text{ 通过结合式(20)}$$

就有

$$\min_{k=1,2,\dots,K} F(\mathbf{w}(k\tau)) - F(\mathbf{w}^*) \leq \varepsilon_0 + \rho h(\tau) \quad (21)$$

2.6 控制算法

2.6.1 近似解

假设本地更新消耗的时间资源 c 和全局聚合消耗的时间资源 b 已知, 并以此计算出 τ 和 T 的值, 我们能够通过 c 和 b 以及其他参数在结合实际情况下, 提出一种能够估计参数并且能动态调整 τ 值的控制算法.

首先, 当 $\eta \leq 1/\beta$ 时, 选择式(15)的上界作为 $F(\mathbf{w}^f) - F(\mathbf{w}^*)$ 的近似值^[16]. 对于全局损失函数 $F(\mathbf{w})$, 其最小值 $F(\mathbf{w}^*)$ 是一个常数, 通过重新将式(7)里的不等式约束排列写做:

$$\min_{\tau, K \in \{1, 2, \dots, k\}} \frac{1}{2\eta\varphi T} + \sqrt{\frac{1}{4\eta^2\varphi^2 T^2} + \frac{\rho h(\tau)}{\eta\varphi\tau}} + \rho h(\tau), \quad K \leq \frac{p}{c\tau + b}, \quad T = K\tau \quad (22)$$

因为 $T = KT_0$, 式(17)中的目标函数随着 T 与 K 减小. 所以对于任意的 τ , K 最优值是 $\min \frac{p}{cT + b}$.

通过忽略舍入运算和替换:

$$T = K\tau \approx \min \frac{p\tau}{c\tau + b} = 1/\max \frac{c\tau + b}{p\tau}$$

得出:

$$G(\tau) = \frac{\max \frac{c\tau + b}{p\tau}}{2\eta\varphi} + \sqrt{\frac{\left(\max \frac{c\tau + b}{p\tau}\right)^2}{4\eta^2\varphi^2} + \frac{\rho h(\tau)}{\eta\varphi\tau^2}} + \rho h(\tau) \quad (23)$$

据此 τ 的最佳解为

$$\tau^* = \underset{\tau \in \{1, 2, \dots, n\}}{\operatorname{argmin}} G(\tau) \quad (24)$$

命题 1 当 $\eta \leq 1/\beta, \rho > 0, \beta > 0, \delta > 0$, 有 $\lim_{p_{\min} \rightarrow \infty} \tau^* = 1$.

命题 2 当 $\eta \leq 1/\beta, \rho > 0, \beta > 0, \delta > 0$, 存在一个有限值 τ_0 , 使得 $\tau^* \leq \tau_0, \tau_0 = \max\{\bullet\}$.

根据命题 2 可以在有限制之下计算 $G(\tau)$, 并找到

使 $G(\tau)$ 最小化的 τ^* . 其中, $v = \operatorname{argmax} \frac{b}{p}$, $V = \operatorname{argmax} \frac{c}{p}$,

$$B = \eta\beta + 1, C_1 = 2\eta\phi p'_v, C_2 = 4\eta^2\phi^2 p_v'^2.$$

2.6.2 自适应边缘联邦学习算法

根据上述理论,此节给出了完整控制算法,每次全局聚合都会重新计算 τ^* ,边缘节点进行本地更新同时记录消耗的时间与能量,在进行聚合器学习时计算经济成本,聚合器将初始模型参数 $\mathbf{w}(0)$ 发送至参与方的边缘节点处,算法 1 给出了聚合器实施过程.

边缘节点实施过程如算法 2 所示.

3 实验

3.1 实验设置

实验为了对算法进行相关性能评估,选择在 5 个节点的真实网络环境以及 5~100 个节点的虚拟环境进行,为了表示异构的边缘计算环境,使用了五台同型号的笔记本电脑将设备用 WiFi 连接在实验室,节点分别训练自身的本地数据集,聚合器则是在一台台式电脑上进行操作,表 1 给出了具体硬件参数.

(1)资源:模拟环境中模拟资源消耗根据高斯分布产生, c_m 和 b_m 分别对应于节点本地模型更新和中心方聚合器.

(2)基线:中心梯度下降、固定 τ 值的传统联邦学习.

(3)模型与数据集:模型采用包括平方支持向量机与卷积神经网络.数据集:PlantVillage^[17](农作物病害图像数据集)数据集,包含 54 306 张图像,其中,40 000 个图像用于训练,剩余 14 306 个图像用于模型测试.

(4)不同节点数据集分布情况设置一共分为四种.案例 1:节点会接受到被随机分布的数据样本;案例 2:节点的数据标签一样,但节点之间的信息不同;案例 3:节点有全部数据集;案例 4:将节点分为两部分,前一部分节点接受随机分配的标签数据,后部分节点与案例 2 相同.

(5)训练与控制参数:时间设置 15 s,搜索范围参数 $\gamma = 10$, τ 值最大为 100,对于支持向量机 $\phi = 0.025$,卷积神经网络 $\phi = 5 \times 10^{-5}$, $\eta = 0.01$, t_d 约为 4.9×10^{-3} s, t_j 为 0.01 J, e_d 为 0.16 s, e_j 为 0.02 J,根据 S 的定义,统一时间与能量的成本,1 s 为 $(1-S)$ 元,1 J 为 S 元,设置 $S = 0.5$ 元.

3.2 实验结果

实验在现实的环境中训练了支持向量机、卷积神经网络两种模型,并且在图 1~4 与基线进行了损失函数值和准确率比较.

卷积神经网络和支持向量机使用随机梯度下降法进行模型训练,在图 1 与图 3 中,横坐标表示本地更新 τ

算法 1 中心方基于边缘计算的环境监测自适应联邦学习

输入:时间资源预算 P ,控制参数 ϕ ,搜索范围参数 γ ,最大的 τ 值;

输出:最终模型 \mathbf{w}^f .

```

1  初始化  $\tau^* \leftarrow 1, t \leftarrow 0, s \leftarrow 0$ ;
2  初始化  $\mathbf{w}(0)$  为常数或随机量;
3  初始化  $\mathbf{w}^f \leftarrow \mathbf{w}(0)$ ;
4  REPEAT
5      聚合器发送  $\mathbf{w}(t)$  和  $\tau^*$  到边缘节点
6       $t_0 \leftarrow t$ ; // 将  $\mathbf{w}(t)$  传输最后的索引迭代保存
7       $t \leftarrow t + \tau^*$ ; // 在  $\tau$  迭代后再全局聚合
8      从  $i$  节点处收集  $e_d, e_j, t_d, t_j$ ; 计算  $\mathbf{w}_i(t), m_{all}$ ;
      根据式(5)计算  $\mathbf{w}(t)$ 
9  IF  $t_0 > 0$  THEN
10     从各  $i$  节点处收集  $F_i(\mathbf{w}(t_0)), \rho_i, \beta_i$ ;
       $\nabla F_i(\mathbf{w}(t_0))$ 
11     根据式(2)计算  $F(\mathbf{w}(t_0))$ 
12     IF  $F(\mathbf{w}(t_0)) < F(\mathbf{w}^f)$  THEN
13          $\mathbf{w}^f \leftarrow \mathbf{w}(t_0)$ ;
14     IF STOP FLAG 已设置 THEN
15         BREAK;
16     估计  $\rho \leftarrow \frac{1}{D} \sum_{i=1}^N D_i \rho_i$ ;
17     估计  $\beta \leftarrow \frac{1}{D} \sum_{i=1}^N D_i \beta_i$ ;
18     根据下面公式计算
       $\nabla F(\mathbf{w}(t_0)) \leftarrow \frac{1}{D} \sum_{i=1}^N D_i \nabla F_i(\mathbf{w}(t_0))$  估计各  $i$  节点的  $\delta_i$ ;
       $\delta_i \leftarrow \nabla F_i(\mathbf{w}(t_0)) - \nabla F(\mathbf{w}(t_0))$  从中估算  $\delta \leftarrow \frac{1}{D} \sum_{i=1}^N D_i \delta_i$ 
19     在  $[1, \tau_m]$  内线性搜索,根据式(24)计算  $\tau$  值, 设置
       $\tau \leftarrow \min\{\gamma\tau^*, \tau_{max}\}$ ;
20     使用从所有  $i$  节点接收的  $c$  和聚合器处的本地测量值
      估计资源消耗  $c, b$ 
21      $s \leftarrow s + c\tau + b$ 
22     IF  $s + c(\tau + 1) + 2b \geq P$  THEN
23         将  $\tau^*$  减小到可能值,以便剩余迭代的估计资源消耗时
      间预算  $P$  内,设置 STOP 标志
24     向所有边缘节点发送  $\mathbf{w}(t)$ ;
25     从每个  $i$  节点接收  $F_i(\mathbf{w}(t))$ ;
26     根据式(2)计算  $F_i(\mathbf{w}(t))$ ;
27     IF  $F(\mathbf{w}(t)) < F(\mathbf{w}^f)$  THEN
28          $\mathbf{w}^f \leftarrow \mathbf{w}(t)$ .
```

值,纵坐标表示基于训练数据的损失函数值,在图 2 与图 4 中,横坐标表示本地更新 τ 值,纵坐标表示基于测试数据的分类准确率,实验中设置本地更新 τ 值从 1 至 100,并将四种案例依次在不同的固定 τ 值里训练,并得到不同 τ 值训练下获得的损失函数值与准确率.

算法 2 节点基于边缘计算的环境监测自适应联邦学习

输入: $w(t)$ 与 τ^* ;

输出: $F_i(w(t))$.

```

1  初始化  $t \leftarrow 0$ ;
2  REPEAT
3    从聚合器接收  $w(t)$  和新  $\tau^*$ , 设置  $w_i(t) \leftarrow w(t)$ 
4     $t_0 \leftarrow t$ ; // 保存  $w(t)$  最后传输迭代索引
5    IF  $t > 0$  THEN
6      估计  $\rho_i \leftarrow F_i(w_i(t)) - F_i(w(t)) / w_i(t) - w(t)$ 
7       $\beta_i \leftarrow \nabla F_i(w_i(t)) - \nabla F_i(w(t)) / w_i(t) - w(t)$ 
8    FOR  $\mu = 1, 2, \dots, \tau^*$  DO
9       $t \leftarrow t + 1$ ; // 进行下轮迭代
10     根据式(4)执行本地更新获得  $w_i(t)$ 
11     IF  $\mu < \tau^*$  THEN
12       设置  $\tilde{w}_i(t) \leftarrow w_i(t)$ ;
13     估计  $i$  节点处进行局部更新消耗时间资源  $c_m$ , 记录每次迭代的时间消耗  $t_d$ , 能量损耗  $t_d$ 
14     发送  $w_i(t), c, i$ , 至聚合器
15     IF  $t_0 > 0$  THEN
16       发送  $\rho_i, \beta_i, F_i(w(t_0)), \nabla F_i(w(t_0))$  至聚合器
17   UNTIL 接收到 STOP 标志
18   从聚合器接收  $w(t)$ ;
19   发送  $F_i(w(t))$  至聚合器
    
```

表 1 边缘节点与聚合器硬件参数

平台	硬件	计算资源	内存	硬盘
边缘节点	笔记本电脑	CPU:R5-3500H GPU:RTX3050	8 GB	512 GB
聚合器	PC 机	CPU:i5-8400 GPU:RTX2060	16 GB	1 TB

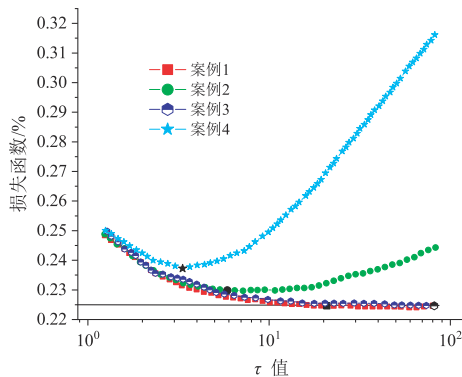


图 1 支持向量机(随机梯度下降)损失函数

算法(具有自适应 τ 的损失函数值)与基线方法进行了比较,并比较了 SVM 和 CNN 分类器的分类精度.在每种情况下,所提出的方法只有一个数据点(由图中黑标记表示),因为在这种情况下 τ 的值是自适应的,

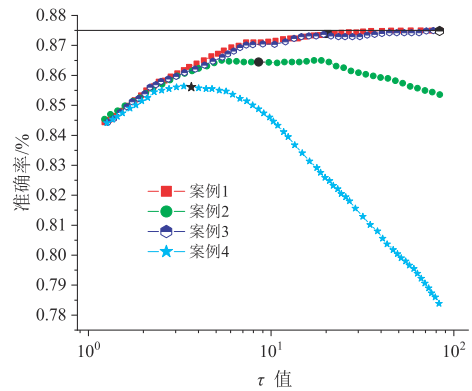


图 2 支持向量机(随机梯度下降)准确率

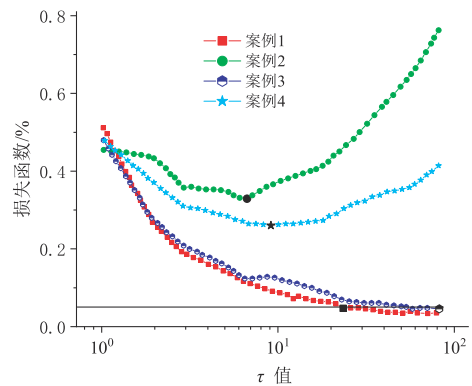


图 3 卷积神经网络随机梯度下降损失函数

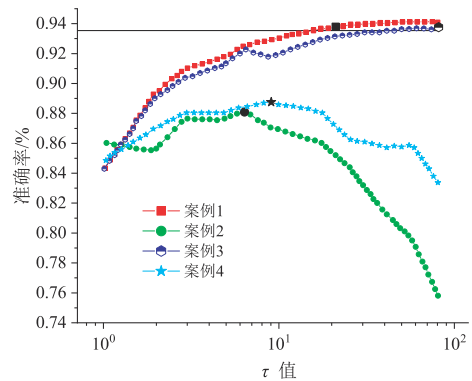


图 4 卷积神经网络(随机梯度下降)准确率

并且标记位置显示了具有相应损失或准确性的平均 τ^* .为了便于比较,在不同的 τ 值之间显示了一条平坦的线.

根据支持向量机和卷积神经网络两种模型的实验,准确率和损失函数值对于不同案例其 τ 值处于不同的位置,因此,固定 τ 值的情况并不能很好适应于全部情况,但是在固定 τ 值为 10 的情况在四种案例都能够获得相对最优的结果,因为在迭代次数较少的情况下,模型的训练过程更容易收敛到局部最优解,而在迭代次数增加后,模型容易陷入过拟合,导致准确率下降.此外,当迭代次数增加时,由于设备资源的限制,节点之

间的通信和计算时间也会增加,从而导致模型性能下降.因此,通过固定迭代次数为10,可以得到相对最优的结果.

在案例2中,由于存在较大的数据集不平衡性,某些节点可能没有足够的数据来提供有意义的更新.这导致一些节点贡献较小,从而影响了全局模型的性能.类似地,在案例4中,一些节点可能拥有更多的数据而另一些节点则拥有较少的数据.这种数据不平衡性也可能导致模型的性能下降.接下来将评估时间预算在自适应 τ^* 值下的表现情况,并对比时间预算下自适应 τ^* 值与固定 τ 值为10训练模型的损失函数值与准确率,由于卷积神经网络模型的高度复杂性,实验重点在支持向量机继续展开,如图5所示.

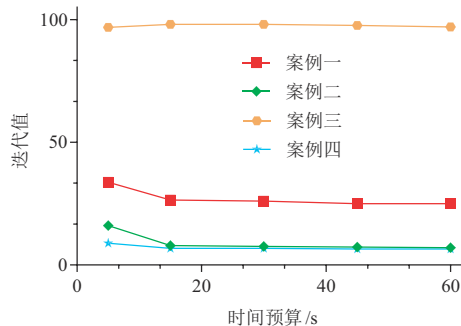


图5 总时间预算

100个节点的虚拟环境中进行自适应迭代 τ^* 值训练模型实验,图中横坐标为总时间预算,纵坐标为最佳迭代 τ^* 值,在节点拥有的数据集不为案例3时, τ^* 值随着时间预算的不断减少,并且随着时间预算的不断增多 τ^* 值基本在10以下,而拥有案例3的节点,因为具有全部数据集,随着时间预算增多 τ^* 值维持在95左右范围.因此,在总时间预算较小的情况下,全局聚合间隔 τ^* 较大,因为需要更多的局部训练轮数来达到相同的分类准确度.相反,在总时间预算较大的情况下, τ^* 减小,因为较少的局部训练轮数已经足够达到相同的分类准确度.

为了全面评估环境监测自适应联邦学习算法在边缘计算系统中的性能,数据集选择了PlantVillage,并采用了上述提到的Case2的数据集划分方式.这种划分方式能更好地模拟边缘设备受到资源约束的情况.在本研究中,图6、图7给出了本文算法与几种经典联邦学习算法进行的对比实验.这些算法包括联邦平均算法(Federated Averaging, FedAvg)、分层边缘联邦学习算法(Hierarchical Federated Edge Learning, HFEL)^[18]以及卷积神经网络联邦学习算法(Convolutional Neural Networks-Federated Learning, CNN-FL)^[19].

通过对比实验的结果,环境监测自适应联邦学习

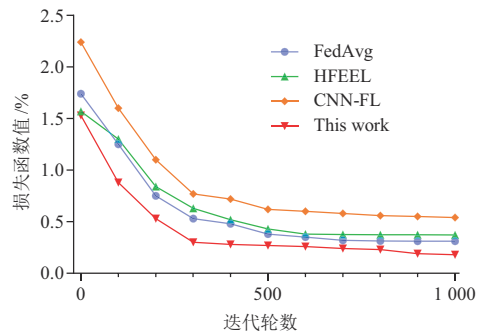


图6 模型损失函数值对比

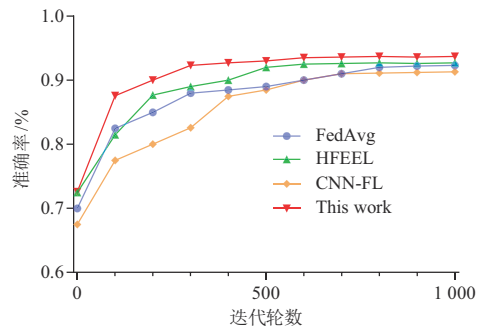


图7 模型准确率对比

算法相较于传统的FedAvg算法、CNN-FL算法以及HFEEL算法,在模型收敛速度方面表现更优越.这也证明本文算法能够更快地训练模型并达到较高的准确性.环境监测自适应联邦学习算法在资源受限的边缘计算系统中利用了资源分配和通信开销的优化策略,确保每一个监测设备的训练任务得到理想的分配.通过调整局部迭代轮数和通信频率等参数,算法能够最大程度地提高模型的收敛速度.面对环境监测数据的实时性、大规模数据处理和离线环境监测等要求时,其他算法无法很好地适应这些挑战.而本文算法通过结合边缘计算系统和自适应机制,能够更好地满足环境监测的需求,从而加速模型训练的过程.实验结果也进一步验证了算法在环境监测领域中的优势,并为其在实际应用中的可行性提供了支持.

图8为经济成本对比结果,可以观察到在100个节点下,四种训练模型在四种案例下的经济成本估计情况.基于这些结果,可以得出以下结论,首先,环境监测自适应联邦学习算法在经济成本方面展现出明显的优势.该算法在资源分配和任务调度过程中充分考虑了能量消耗的影响.通过动态分配任务和资源,环境监测自适应联邦学习算法能够更好地利用边缘设备的能源,最小化能量消耗.相比之下,传统的FedAvg、CNN-FL和HFEEL等算法没有针对能量消耗进行优化,因此在经济成本方面存在一定的局限性.其次,环境监测自适应联邦学习算法通过灵活的资源管理策略,能够有

效降低经济成本. 算法根据不同案例的特点和要求进行资源分配, 以最大限度地减少经济成本. 这使得算法能够在经济有限的条件下获得更好的模型更新和学习效果.

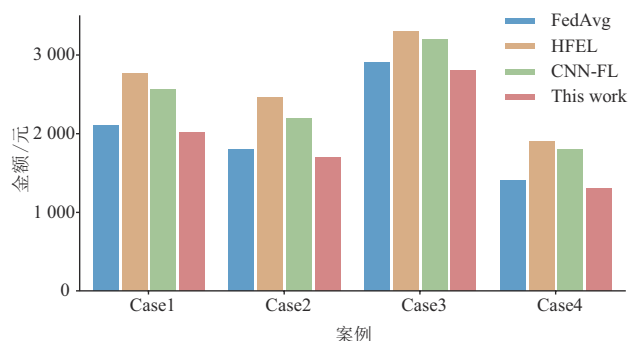


图8 经济成本

4 总结

针对环境监测领域日益增长的智能设备和大规模数据处理需求, 本文在综合资源约束、设备异构性、数据分布的非独立同分布等问题的基础上提出了一种基于边缘计算的环境监测自适应联邦学习算法, 具有很高的实用价值, 由于本文重点在环境监测领域, 未来的研究可以在不同领域下验证和扩展算法的适用性.

参考文献

- [1] MONTRUCCHIO B, GIUSTO E, VAKILI M G, et al. A densely-deployed, high sampling rate, open-source air pollution monitoring WSN[J]. *IEEE Transactions on Vehicular Technology*, 2020, 69(12): 15786-15799.
- [2] ZHU B T, BEDEER E, NGUYEN H H, et al. UAV trajectory planning for AoI-minimal data collection in UAV-aided IoT networks by transformer[J]. *IEEE Transactions on Wireless Communications*, 2023, 22(2): 1343-1358.
- [3] SABARI M, ASWINTH P, KARTHIK T, et al. Water quality monitoring system based on IoT[C]//2020 5th International Conference on Devices, Circuits and Systems (ICDCS). Piscataway: IEEE, 2020: 279-282.
- [4] 刘艳, 王田, 彭绍亮, 等. 基于边缘的联邦学习模型清洗和设备聚类方法[J]. *计算机学报*, 2021, 44(12): 2515-2528.
LIU Y, WANG T, PENG S L, et al. Edge-based model cleaning and device clustering in federated learning[J]. *Chinese Journal of Computers*, 2021, 44(12): 2515-2528. (in Chinese)
- [5] LIU Y, HUANG A B, LUO Y, et al. FedVision: An online visual object detection platform powered by federated learning[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(8): 13172-13179.
- [6] YANG Z H, CHEN M Z, SAAD W, et al. Energy efficient federated learning over wireless communication networks[J]. *IEEE Transactions on Wireless Communications*, 2021, 20(3): 1935-1949.
- [7] LIU Y, NIE J T, LI X D, et al. Federated learning in the sky: Aerial-ground air quality sensing framework with UAV swarms[J]. *IEEE Internet of Things Journal*, 2021, 8(12): 9827-9837.
- [8] DEY S, PAL S. Federated learning-based air quality prediction for smart cities using BGRU model[C]//Proceedings of the 28th Annual International Conference on Mobile Computing and Networking. New York: ACM, 2022: 871-873.
- [9] T M, MAKKITHAYA K, G N V. A federated learning-based crop yield prediction for agricultural production risk management[C]//2022 IEEE Delhi Section Conference (DELCON). Piscataway: IEEE, 2022: 1-7.
- [10] SUN T, TANG K, LI D S. Gradient descent learning with floats[J]. *IEEE Transactions on Cybernetics*, 2022, 52(3): 1763-1771.
- [11] ABBAS N, ZHANG Y, TAHERKORDI A, et al. Mobile edge computing: A survey[J]. *IEEE Internet of Things Journal*, 2018, 5(1): 450-465.
- [12] MILLS J, HU J, MIN G Y. Communication-efficient federated learning for wireless edge intelligence in IoT[J]. *IEEE Internet of Things Journal*, 2020, 7(7): 5986-5994.
- [13] SULIMAN M, LEITH D J. Gradient information from google GBoard NWP LSTM is sufficient to reconstruct words typed[C]//2022 Cyber Research Conference - Ireland (Cyber-RCI). Piscataway: IEEE, 2023: 1-4.
- [14] YANG M M, TJUAWINATA I, LAM K Y. K-means clustering with local ϵ -privacy for privacy-preserving data analysis[J]. *IEEE Transactions on Information Forensics and Security*, 2022, 17: 2524-2537.
- [15] MAO Y Y, YOU C S, ZHANG J, et al. A survey on mobile edge computing: The communication perspective[J]. *IEEE Communications Surveys & Tutorials*, 2017, 19(4): 2322-2358.
- [16] JIE T. Research on key technologies of advanced electronic measurement sensors based on computer Internet of Things[C]//2022 IEEE 4th International Conference on Civil Aviation Safety and Information Technology (IC-CASIT). Piscataway: IEEE, 2022: 856-860.
- [17] KIRTI, RAJPAL N. Black rot disease detection in grape plant (*vitis vinifera*) using colour based segmentation &

machine learning[C]//2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN). Piscataway: IEEE, 2021: 976-979.

- [18] WEN W L, CHEN Z H, YANG H H, et al. Joint scheduling and resource allocation for hierarchical federated edge learning[J]. IEEE Transactions on Wireless Communications, 2022, 21(8): 5857-5872.
- [19] SALIM M M, PARK J H. Federated learning-based secure electronic health record sharing scheme in medical informatics[J]. IEEE Journal of Biomedical and Health Informatics, 2022, 27(2): 617-624.

作者简介



蒋伟进 男, 1964年7月出生于湖南省益阳市. 现为湖南工商大学计算机学院二级教授、硕士生导师. 主要研究方向为联邦学习、群智感知、边缘计算、社会计算.
E-mail: jwjnudt@163.com



韩裕清(通讯作者) 男, 2000年11月出生于湖南省长沙市. 现为湖南工商大学计算机学院硕士研究生. 主要研究方向为联邦学习、边缘计算、群智感知.
E-mail: 897735614@qq.com



吴玉庭 女, 1998年4月出生于湖南省益阳市. 现为湖南工商大学计算机学院硕士研究生. 主要研究方向为联邦学习、群智感知.
E-mail: 1321224262@qq.com



周为 男, 2000年5月出生于湖南省益阳市. 现为湖南工商大学计算机学院硕士研究生. 主要研究方向为联邦学习、群智感知.
E-mail: 1216330671@qq.com



陈艺琳 女, 2000年9月出生于河南省许昌市. 现为湖南工商大学计算机学院硕士研究生. 主要研究方向为联邦学习、信息安全.
E-mail: 1986746095@qq.com



王海娟 女, 2000年5月出生于江西省九江市. 现为湖南工商大学前沿交叉学院硕士研究生. 主要研究方向为联邦学习、群智感知.
E-mail: 2502560411@qq.com