

边缘场景下动态权重的联邦学习优化方法

程帆, 王瑞锦, 张凤荔

引用本文

程帆, 王瑞锦, 张凤荔. 边缘场景下动态权重的联邦学习优化方法[J]. 计算机科学, 2022, 49(12): 53-58.

CHENG Fan, WANG Rui-jin, ZHANG Feng-li. [Federated Learning Optimization Method for Dynamic Weights in Edge Scenarios](#) [J]. Computer Science, 2022, 49(12): 53-58.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[融合多特征的属性异质网络嵌入方法](#)

Method of Attributed Heterogeneous Network Embedding with Multiple Features

计算机科学, 2022, 49(12): 146-154. <https://doi.org/10.11896/jsjcx.211200082>

[基于联邦学习的暖通空调系统故障检测与诊断](#)

Fault Detection and Diagnosis of HVAC System Based on Federated Learning

计算机科学, 2022, 49(12): 74-80. <https://doi.org/10.11896/jsjcx.220700280>

[基于联邦学习的Gamma回归算法](#)

FL-GRM: Gamma Regression Algorithm Based on Federated Learning

计算机科学, 2022, 49(12): 66-73. <https://doi.org/10.11896/jsjcx.220600034>

[基于联邦学习的车联网多维资源动态分配算法](#)

Multi-dimensional Resource Dynamic Allocation Algorithm for Internet of Vehicles Based on Federated Learning

计算机科学, 2022, 49(12): 59-65. <https://doi.org/10.11896/jsjcx.211000123>

[联邦学习激励机制研究综述](#)

Survey of Incentive Mechanism for Federated Learning

计算机科学, 2022, 49(12): 46-52. <https://doi.org/10.11896/jsjcx.220500272>

边缘场景下动态权重的联邦学习优化方法

程 帆 王瑞锦 张凤荔

电子科技大学信息与软件工程学院 成都 610054

(fancheng@std.uestc.edu.cn)

摘 要 边缘计算(Edge Computing)作为一种新的计算范式,在网络边缘提供计算服务,相比传统的云计算模式,它具有高可信、低延迟等特点,在各行各业中有着广阔的应用前景,但在隐私保护和数据处理上仍存在一些问題。而联邦学习作为一种分布式的机器学习技术,能很好地解决边缘计算场景下数据分布不一致和数据隐私问题,但仍面临设备异构、数据异质及通信方面的挑战,如模型偏移、收敛效果差、部分设备计算结果丢失等问題。为解决上述问題,提出动态权重的联邦学习优化算法(FedDw)。该算法关注设备的服务质量,减少训练速度不一致导致部分设备参与带来的异构性影响,并根据服务质量确定在最终模型聚合时的占比,从而确保聚合的结果在复杂的真实情况下更具有鲁棒性。在 10 个地区气象站的真实数据集上与 FedProx 和 Scaffold 这两种典型的联邦学习算法进行了对比,实验结果表明 FedDw 算法具有更好的综合性能。

关键词 联邦学习;边缘计算;风能预测;设备异构;动态权重

中图法分类号 TP391

Federated Learning Optimization Method for Dynamic Weights in Edge Scenarios

CHENG Fan, WANG Rui-jin and ZHANG Feng-li

School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China

Abstract As a new computing paradigm, edge computing provides computing and storage services at the edge of the network compared to traditional cloud computing model. It has the characteristics of high reliability and low latency. However, there are still some problems in privacy protection and data processing. As a distributed machine learning model, federated learning can well solve the problems of inconsistent data distribution and data privacy in edge computing scenarios, but it still faces challenges in equipment heterogeneity, data heterogeneity and communication, such as model offset, the convergence effect is poor, and the calculation results of some devices are lost. In order to solve the above problems, a federated learning optimization algorithm with dynamic weights(FedDw) is proposed, which focuses on the service quality of the equipment, reduces the heterogeneous impact caused by the participation of some equipments due to inconsistent training speed, and determines the proportion in the final model aggregation according to the service quality, so as to ensure that the aggregation results are more robust in complex real situations. Through experiments, the two excellent federated learning algorithms, FedProx and Scaffold, are compared on the real data sets of 10 regional weather stations. The results show that the FedDw algorithm has better comprehensive performance.

Keywords Federal learning, Edge computing, Wind energy forecasting, Equipment heterogeneity, Dynamic weights

1 引言

近年来,随着物联网(Internet of Things, IoT)的迅速发展,云计算已无法满足海量边缘数据爆炸式增长的需求^[1-2],边缘计算以其更接近物联网终端设备、能在边缘提供计算

服务、高可信、低延迟的优点,逐渐在各行各业得到了广泛的应用,如智慧城市、工业物联网、智能家居等^[3-6]。但边缘计算在数据的预处理和隐私保护方面仍面临巨大难题^[7],以风力发电场景为例:全球风能理事会(GWEC)表示每年风电的新增储量超过 50 GW,但是由于风力发电的

到稿日期:2022-07-15 返修日期:2022-09-08

基金项目:国家自然科学基金(62271128);四川省科技计划重点研发项目(2022ZDZX0004, 23ZDYF0706, 23ZDYF0085, 2022YFG0212, 2021YFS0391, 2021YFG0027, 2020YFG0475, 2019YJ0543)

This work was supported by the National Natural Science Foundation of China(62271128) and Key R & D Projects of Sichuan Science and Technology Program(2022ZDZX0004, 23ZDYF0706, 23ZDYF0085, 2022YFG0212, 2021YFS0391, 2021YFG0027, 2020YFG0475, 2019YJ0543).

通信作者:王瑞锦(ruijinwang@uestc.edu.cn)

不稳定性和低效率性,供电需求和电力供应经常不匹配,从而造成电力浪费和设备故障情况,因此风能预测对风力发电的发展有着重要意义。传统的风能预测的方法有基于数值天气预报(NWP)的预测、时间序列预测、基于人工神经网络预测等,但由于各地的风力数据通常都是质量很差的数据且数据分布较为分散,同时这类数据大多属于隐私信息,容易形成“数据孤岛”,导致传统方法在实际中对风能预测效果并不是很理想。因此,考虑通过联邦学习对数据进行增强,以辅助其他的风能预测方法对风力发电问题进行进一步优化^[8]。

近年来,联邦学习(Federated Learning)作为分布式的机器学习模型逐渐成为研究的热点^[9-11]。和传统的集中式机器学习模型不同,联邦学习的数据集不是存储在中央服务器上,而是分布在大量的客户端上,这些客户端可以是手机、监控摄像头、网联汽车等拥有计算能力的终端设备。当终端设备产生大量数据时,可以直接通过本地的模型进行训练,而不需要通过网络传输数据给中央服务器,从而能够在提升模型训练效率的同时保护用户的隐私。

传统的集中式机器学习模型通信成本相对较小,计算成本则占主导地位,通常依靠使用 GPU 来降低相关成本,而在联邦学习中通信成本则占主体地位。总结以往的研究,主要存在以下 3 方面的挑战:

- (1)通信难题:服务器与客户端间网络连接相对较慢。
- (2)设备异构:在给定时间内只有少部分客户端能够用于训练,并且不同设备的计算能力可能不同。
- (3)数据异质:不同客户端间数据分布不一致^[12]。

针对通信难题,目前已有通过模型压缩和模型蒸馏来降低链路通信成本的方法;针对设备异构,已有研究通过元学习来实现不同设备的个性化服务;针对数据异质,已有研究通过在本地模型的目标函数和训练轮数中引入正则化项来控制偏移量,使得本地模型的优化方向不会偏离全局模型太远。而在已有的联邦学习优化算法中,FedAvg^[13]作为最早提出的优化算法,其将移动设备分散数据的训练问题作为重要的研究方向,但对数据异质和设备异构没有很好的解决办法;FedProx^[14]和 FdeNova^[15]通过在本地训练中加入正则项来控制不同本地模型的优化方向,虽没有引入额外不可接受的开销,但需要本地模型在每一次训练时对正则化项的权重参数进行调优;Scaffold^[16]则假设每个本地模型“可视化”,通过重复利用之前计算的梯度来更新局部控制变量,同时,每次通信不仅要传输模型的参数,还要传输控制变量的参数,增加了单次通信的代价。上述算法只针对数据异质或设备异构进行优化,并没有同时考虑两种情况,过于理想化。

为了解决风能预测中数据分布不一致和数据隐私问题,本文在已有联邦学习框架的基础上提出了一种新的联邦学习算法。本文的主要贡献如下:

提出了动态权重的联邦学习优化算法。联邦学习为了保护用户的隐私,不同终端设备的本地数据集为不代表总体分

布的非独立同分布数据(Non-IID),而针对这种 Non-IID 的数据,以往的联邦优化算法在通信效率和计算能力上仍有缺陷。本文提出的算法将注意力机制^[17]和联邦学习结合,增加梯度对目标域有益的节点的权重,限制梯度对目标域有害的节点权重。根据设备返回服务器的计算能力和训练时间值来确定每个设备在最终聚合时所占的权重,保证全局模型和本地模型的协同性,在确保模型精准度的基础上减小通信消耗。

我们获取了不同地区 10 个城市的真实用电负荷数据和风功率数据,并通过动态改变设备的计算能力来模拟异构环境进行相关实验,与现有联邦学习算法进行对比,证明了所提算法的可行性。

2 相关工作

2.1 注意力机制

注意力机制已被广泛应用在图像识别和自然语言处理中,从本质上来说类似于人类的选择性视觉注意力,将处理对象聚焦在关键信息中。在和联邦学习的结合中,已有研究考虑将本地训练的 CNN 模型替换成 Transformer 结构,从而延长模型的记忆能力,避免灾难性遗忘(Catastrophic Forgetting)^[18-19]。而本文主要考虑借助注意力机制的思想,即关注本地模型的服务质量,增加服务质量高的设备的权重,降低服务质量低的设备的权重。

2.2 已有联邦学习算法

联邦学习的提出是为了更好地保护数据的隐私,但同时也要确保建模效果和将整个数据集放在一起建模的效果相差不多。但是,我们注意到数据的异构性即 Non-IID 问题会影响最终聚合的模型,而针对联邦学习这种数据 Non-IID 问题,已有研究在 FedAvg 的基础上,对模型更新和模型聚合进行改进,通过研究特征标签分布偏差、数量偏差来提高模型的鲁棒性和优化效率;同时也有研究采用个性化联邦学习^[20],针对不同用户对模型性能的要求不一致来进行优化;此外也有部分文献在数据层结合多类型数据进行研究,通过对数据的增强来解决 Non-IID 问题^[15]。但是上述算法很难在通信成本和计算成本之间达到一个平衡,不能动态地调整模型更新方向,因此本文提出利用注意力机制的动态权重的联邦学习优化算法(Federated Dynamic weights Algorithm, FedDw)。该算法能动态地调整模型更新方向,保证可靠的模型上传,能够很好地容忍设备异构性,并对网络中的离线设备具有很好的鲁棒性。实验表明, FedDw 相比传统联邦学习算法在模型准确率和收敛速度上有明显优势。

3 动态权重-联邦学习优化算法

3.1 问题描述

联邦学习的优化目标是最小化所有样本的平均损失,如式(1)所示:

$$f(\omega) = \sum_{k=1}^K \frac{n_k}{n} F_k(\omega) \quad \text{where} \quad (1)$$

$$F_k(\omega) = \frac{1}{n_k} \sum_{i \in P_k} f_i(\omega)$$

其中, $F_k(\omega)$ 表示第 k 个客户端所有数据的平均损失, $f(\omega)$

表示当前参数下所有客户的加权平均损失,即目标函数。

如果所有 P_k (第 k 个客户端的数据)都是通过随机均匀地将训练样本分布在客户端上形成的,即数据都是独立同分布的,那么每一个 $F_k(w)$ 的期望都是 $f(w)$,本地训练得到的模型和全局聚合的模型偏差不大,如图1所示。

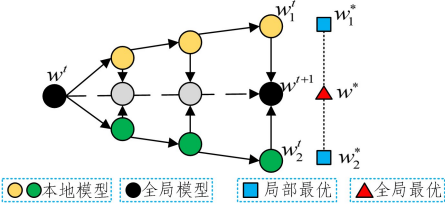


图1 IID数据产生的模型偏移
Fig. 1 Model offset from IID data

不同设备的数据往往为非独立同分布,当服务器初始化全局参数发放给每个客户端,每个客户端根据各自的本地数据训练相同的轮数,由于数据的异质性和设备的异构性可能产生模型偏移,如图2所示,如果按照以往的随机梯度下降更新,并在本地训练结束后对模型取均值,得到的全局模型并不会在最优解的位置。

本文通过以下方法来解决上述问题:

- (1)数据异质方面:在本地设备优化目标中添加余项,防止本地过优化,偏离全局优化方向。
- (2)设备异构方面:引入注意力机制思想,关注设备的服务质量,增加服务质量高的设备的权重,降低服务质量低的设备的权重,3.2节对此作了详细介绍。

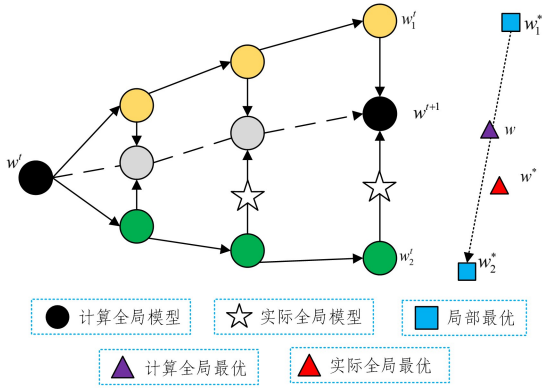


图2 Non-IID数据产生的模型偏移
Fig. 2 Model offset from Non-IID data

在上述理论方法的基础上,我们提出了动态权重的联邦学习优化算法(FedDw),算法的整体架构如图3所示。

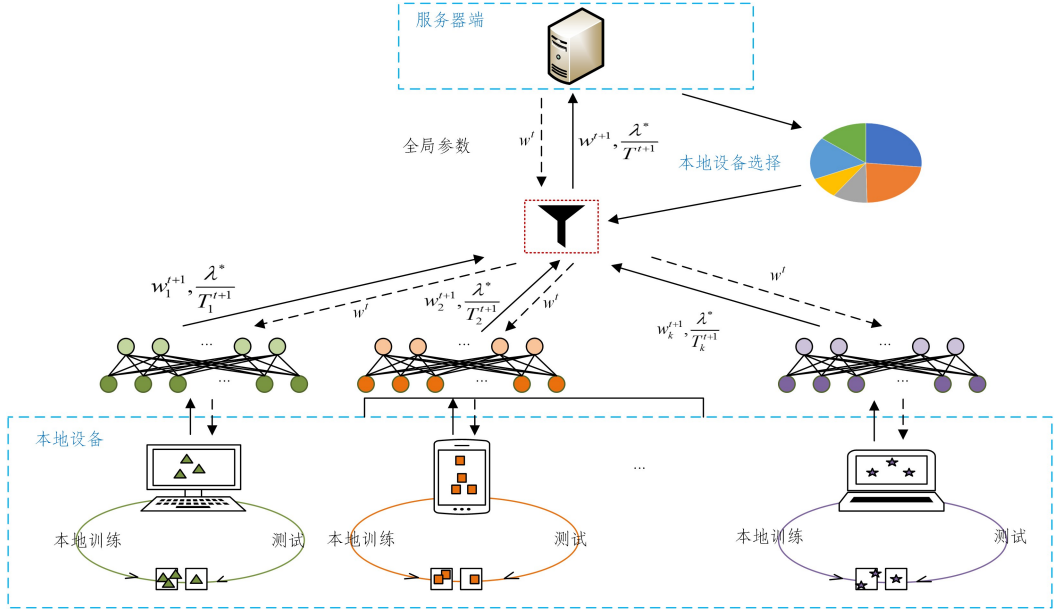


图3 FedDw架构图
Fig. 3 FedDw architecture diagram

3.2 设备权重的动态设置

由于不同设备在计算能力上存在差异,且通信效率会受到时间的影响,训练相同的轮数时,很多设备无法按时完成训

练,从而导致全局模型聚合的效果较差。因此,本文借助注意力机制的思想^[21],关注设备的服务质量 E_k ,根据设备 k 在第 t 轮的计算能力,来设置第 t 轮该设备在最终聚合时的权重。



图4 FedDw与传统联邦学习梯度更新方式对比

Fig. 4 Comparison between FedDw and traditional federated learning gradient update methods

传统的联邦学习中^[22],不同的本地设备因计算能力的差异,在计算同一批次大小的样本数据时,每一轮的时间通常由计算最慢的设备来确定,当有一个设备出现异常或者掉队,其他设备的时间开销可能会同时增大,如图 4 所示。而 FedDw 算法架构在每次全局模型更新中会给不同本地设备在每一轮设置相同的迭代次数而不是相同的时间,将设备 k 在第 t 轮的时间设为 T_k^t ,权重设为 $\frac{\lambda^*}{T_k^t}$,其中 λ^* 为设备在该轮中实时的计算能力。当超过某一固定值 T_{final} 时,认为该设备本轮任务丢失,权重为 0,即当该设备因网络或硬件设施出现异常时,不会对全局模型参数产生影响,减缓模型收敛的速度。当本轮本地设备训练完成时,向服务器回传模型参数 w_k^t 及权重 $\frac{\lambda^*}{T_k^t}$,服务器根据权重聚合模型参数得到新的全局模型 w^{t+1} 。由于设备在每轮的训练时间不同,权重值的设置也会随之动态更新,在一定程度上减轻了设备异构带来的影响。

3.3 动态权重的联邦学习优化算法 FedDw

针对联邦学习中数据的 Non-IID 和不同设备计算能力的差异,我们提出了动态权重的优化算法 FedDw,它在整体框架上类似于 FedAvg,在每一轮训练中选择参与训练的设备集合,进行本地的局部更新,然后在服务器端对这些更新进行聚合形成全局更新。但是在此基础上,我们做了以下简单但关键的改进。

改进一 设备的异构性通过计算能力来模拟。数据的异质性可以通过不同地区的数据来实现,设备的异质性则需要通过分配计算能力的不同来模拟,给予设备不同的计算能力 λ^* ,并在每一轮通过高斯分布的方式,动态地更新不同设备的计算能力。其中平均值 μ 在区间 $[0,1]$ 随机初始化,标准偏差 σ 在区间 $[0.25\mu, 0.5\mu]$ 随机初始化。

改进二 定义不同本地设备的权重。通过训练相同轮数,对比不同设备 k 所表现出的性能(所花费时间、网络是否稳定),寻找相对可靠的客户端,并给定权重 $\frac{\lambda^*}{T_k}$,在最终全局模型聚合时根据权重占比来定义传输的本地模型参数占比,不再是原来的 $w' = (w_1' + \dots + w_i')/i$,而是 $w' = (w_1' * \frac{\lambda^*}{T_1} + \dots + w_i' * \frac{\lambda^*}{T_i})/i$ 。

改进三 重新定义本地更新的目标函数。设备 k 在本地训练时,所需最小化目标函数更新为 $w_k^{t+1} = F_k(w) + \mu \|w - w'\|^2$,在原有 $F_k(w) = \frac{1}{n_k} \sum_{i \in P_k} f_i(w)$ 基础上加入一个余项,即取本地局部优化的参数和服务器传来的全局参数向量的距离差值。通过这种方式,可以确保本地局部更新不会偏离全局优化方向太远,同时能够有效地融合系统异构造成的不同数据间的差异性。

FedDw 算法的主要思想是增加梯度对目标域有益的节点的权重,限制梯度对目标域有害的节点权重,动态调整模型更新方向,加快模型收敛速度。FedDw 算法的伪代码如算法 1 所示。

算法 1 动态权重的联邦学习优化算法 (FedDw)

输入:本地数据集 D ,总客户端数 N ,样本客户端数 M ,通信轮数 T ,初

始全局模型 w^0 ,本地训练轮数 E ,学习率 η ,余项参数 μ ,截止时间 T_{final}

输出:最终全局模型 w^T

- 1. 服务器端:
- 2. 初始化 w^0
- 3. for $t \leftarrow 0$ to $T-1$ do:
- 4. $S_t \leftarrow$ 从 N 个客户端随机采样 M 个客户端
- 5. $n_s \leftarrow \sum_{k \in S_t} |D^k|$ //首先根据概率 p_k 随机选出一批客户端,它们的集合为 S_t ,并为不同的客户端 k 分配对应的数据集 D^k
- 6. for $k \in S_t$ in parallel do:
- 7. 发送全局模型 w_s^t 到客户端 C_k
- 8. $w_k^{t+1} \leftarrow \text{ClientUpdate}(k, w_s^t)$
- 9. $w_s^{t+1} \leftarrow \sum_{k \in S_t} \frac{|D^k|}{n_s} w_k^{t+1} \cdot \frac{\lambda^*}{T_k^{t+1}}$
//根据客户端回传的权重按改进二方式聚合这些参数得到下一轮初始参数
- 10. return w_s^T
- 11. 客户端:
- 12. 动态分配计算能力 λ^* //在每一轮通信中随机生成计算能力 λ^* ,并模拟高斯分布动态更新
- 13. $\text{ClientUpdate}(k, w^t)$:
//本地更新函数,13-23 行,每一个被选中的客户端 k ,根据本地风能数据集对全局模型 w 进行 E 轮训练,若其模型训练总时长 T 小于截止时长 T_{final} ,按改进三方式更新本地模型参数 w_k^{t+1}
- 14. $L(w; b) = \sum_{(x,y) \in b} l(w; x; y)$
- 15. $w_k^t \leftarrow w^t$
- 16. for epoch $\leftarrow 0$ to $E-1$ do:
- 17. for each batch $b = (x, y)$ of D^k do:
- 18. if T_{final} 内完成计算 then
- 19. $T_k^{t+1} \leftarrow$ 设备 k 完成 batch b 的时间
- 20. $F_k(w) \leftarrow w_k^t - \eta \nabla L(w_k^t; b)$
- 21. $w_k^{t+1} \leftarrow F_k(w) + \mu \|w - w^t\|^2$
- 22. else
- 23. $w_k^{t+1} \leftarrow w^t$ //超出 T_{final} ,参数不变
- 24. return $w_k^{t+1}, T_k^{t+1}, \lambda^*$ to the server

4 实验结果及分析

4.1 数据集和模型

本文的数据集源自 2019 年四川省凉山、甘孜、阿坝的 10 个地区,包含气象站从 2019 年 1 月 1 日到 10 月 1 日,每次间隔 1h,在高于水平线 10m 和 100m 处测量的风速、风向数据和发电量。通过固定长度滑动窗口将数据序列化并对数据作归一化处理,将数据划分为两部分,其中 80% 作为训练集,剩下 20% 作为测试集。对风力发电量进行预测,调整用电规模,对提高风能利用率有着重要意义。本文主要聚焦于设备异构和数据异质问题,为了防止模型带来的影响,使用简单的 4 层神经网络,并在与 FedProx 和 Scaffold 进行对比时保证模型的一致性,取 10 次实验的平均值作为对比结果。

4.2 实验设置

FedDw 算法基于 Pytorch 的客户端/服务器端联邦学习框架实现,通过设置相同的迭代次数,根据计算时间来衡量最终聚合时的本地模型参数占比,而不是简单的平均加权。为

了体现 FedDw 算法在数据异质、设备异构环境下的特点和性能,将其与目前表现较好的联邦学习算法 FedProx 和 Scaffold 进行了对比,实验结果将在 4.3 节中讨论。其中,FedProx 算法为了防止设备间数据异质导致每个客户端优化后的模型与初始时服务器分配的全局模型相差过大,在 FedAvg 损失函数的基础上,引入了近端项,确保客户端在本地训练后得到的模型参数 w 不会与初始时的服务器参数 w^0 偏差过大。Scaffold 算法通过在客户端和服务端设置局部和全局控制变量的方式,重复计算之前的梯度来更新局部控制变量,解决数据异质导致的收敛变缓问题。以上两种算法都只对数据异质进行优化,而本文提出的 FedDw 算法同时考虑设备异构和数据异质问题,能够动态调整模型更新方向,保证可靠的模型上传,并且能够很好地容忍设备异构性,对网络中的离线设备具有很好的鲁棒性。由于缺乏计算资源,实验中服务器端的聚合和客户端的本地训练和测试都在同一台具有 NVIDIA 1660Ti 的机器上进行,设备的异质性通过给予机器不同的计算能力来模拟,即限制 GPU 的使用效率。

实验相关参数设置如下,本地训练中,使用的优化器中 SGD 参数设置为:学习率 0.01,动量 0.9,Adam 学习率为 0.01,在默认情况下,本地训练轮数为 10,通信轮数为 5,客户总数为 10,每次以 0.5 的样本率对客户端进行随机抽样,样本输入维度为 28,batch 大小设置为 50。数据异质通过不同地区数据的不同来体现,设备异构通过地区数据对应分配的权重不同来模拟。

4.3 性能比较和结果分析

4.3.1 平均绝对误差和均方根误差

本节使用平均绝对误差(Mean Absolute Error,MAE)和均方根误差(Root Mean Square Error, RMSE)来衡量模型拟合结果的好坏,结果如表 1 所列。

表 1 不同模型的预测对比

Table 1 Comparison of prediction performance of different models

Algorithms	MAE	RMSE
FedProx	0.304	0.345
Scaffold	0.318	0.455
FedDw*	0.150	0.179

注: * 代表本文提出的方法

表 1 列出了在默认情况下 3 种联邦学习算法的对比,其中最优结果用加粗字体表示。在拟合效果方面,FedDw 最优,FedProx 次之。在 3 种联邦学习算法架构的基础上,分别对客户数目、采样率、通信轮数、批次大小、本地训练轮数进行了调整,结果都大致相同,偏差在 5% 以内,因此均使用默认的参数表示算法的结果。我们分析了三者拟合效果都差强人意的原因:1)数据的复杂程度和模拟设备的计算能力差异;2)实验所设计的模型比较简单,这恰巧能对 3 种联邦学习算法有一个很好的区分,从而体现本文所设计的算法在数据异质和设备异构场景下的优点。分析 FedDw 拟合效果优于 FedProx 和 Scaffold 的原因,从直观意义上来讲,风力数据在一定时间段内呈连续状态且在不同日期的同一时间段大致相同,除非出现极端天气;同时设备的计算能力在动态刷新,容易出现训练某一个地区的数据时,出现连接中断和训练时间过长的情况,而 FedDw 正好避免了这方面的问题,有针对性

地对连续性数据和设备的差异性进行了处理。

4.3.2 loss 值下降速率

为了对 3 种算法的性能有一个更直观的认识,本文选取了综合效果最好的地区 4,取地区 4 在 5 次通信的随机选择中均出现 3 次以上的实验结果进行分析,它们的 loss 值对比如图 5 所示。其中 FedDw 在 epoch 1 时 loss 值下降很快,而后呈收敛状态且最终 loss 值远小于其他两种算法。Scaffold 算法在收敛速度和最后的优化效果方面均表现不错,但它在每一轮的结束接近收敛时,以及在下一轮开始的 epoch 1 时 loss 值总是会增大。同时,在进行数据分析时发现,3 种算法在收敛的过程中从前一轮随机选择训练结束后到下一次又被选中开始训练时,loss 值并没有继承,反而有上升的趋势,如图中 FedProx 算法 epoch 10 到 epoch 11、Scaffold 算法 epoch 20 到 epoch 21。一开始,我们认为可能是模型出现了过拟合,但经分析发现,其原因在于实验中设备动态刷新的计算能力:前一次通信中最终聚合的全局模型参数,因为综合了不同地区数据集,在新一轮传输给原来的客户端时,客户端的计算能力发生改变并且全局模型和之前的本地模型存在偏移,导致本地模型需要有个重新适应的过程。

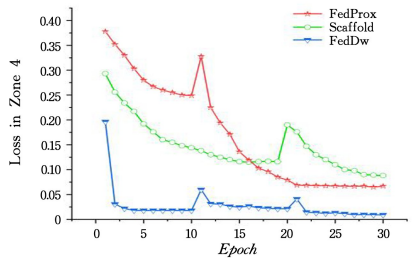


图 5 地区 4 中 3 种算法损失值下降对比

Fig. 5 Comparison of loss values of three algorithms in region 4

4.3.3 平均训练时间

通信效率在评价联邦学习算法的性能上也是一个重要的指标,本文通过 GPU 的使用情况记录了每次通信中本地模型的平均训练时间,如图 6 所示。Scaffold 算法无论是训练开始时的计算时间还是最后收敛时的时间都优于 FedProx 算法和 FedDw 算法,但 FedProx 算法时间在第一轮之后明显缩短,这应该是算法中使用了不精确解的概念。FedDw 算法在时间上表现不佳,原因是其他算法只需要把模型参数回传给服务器端,而 FedDw 还需要把计算时间和本次动态刷新获得的计算能力连同模型参数一起回传,在获得好的预测效果的同时付出了通信上的代价。3 种算法在 10 个地区收敛时的平均时间如图 7 所示,可以看出,当数据的异质性较大时,FedDw 的表现仍相对稳定,具有很好的泛化能力。

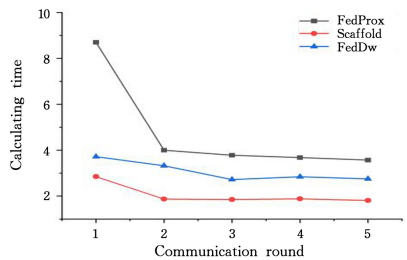


图 6 每次通信中本地模型训练平均时间

Fig. 6 Average local model training time per communication

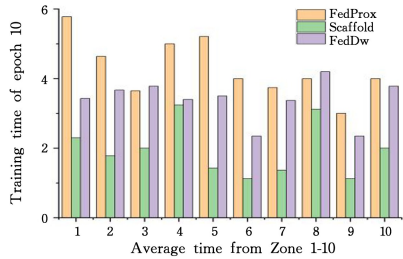


图 7 地区 1—10 模型收敛时平均训练时间

Fig. 7 Average training time when models converge in regions 1—10

结束语 本文提出了一种动态权重的联邦学习优化算法,通过关注设备的服务质量,利用设备服务质量确定模型聚合的方向,得到最终的预测模型。本文算法架构考虑到真实场景中存在的设备异构和数据异质,因此具有一定的普适性,可应用在类似风力发电的边缘场景下,解决不同地区数据保密、系统不同带来的不便性。在未来的工作中,我们将关注设备的选择,减少因随机选择而给实验带来的干扰,挑选出对模型聚合有利的设备,并根据本文的实验现象对收敛过程中产生的波动进行优化,关注算法的通信复杂度,设计效率更高、成本更低、复杂度更低的联邦学习算法。

参 考 文 献

[1] ZHOU Y Z,ZHANG D. Near-ENF cloud computing:opportunities and challenges in the post-cloud computing Era[J]. Chinese Journal of Computers,2019,42(4):677-700.

[2] CUI Y,SONG J,MIAO C C,et al. Mobile cloud computing research progress and trends[J]. Chinese Journal of Computers,2017,40(2):273-295.

[3] ARRIETA A B,DÍAZ-RODRÍGUEZ N,DEL SER J,et al. ExplainableArtificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI[J]. Information Fusion,2020,58:82-115.

[4] SHI W S,SUN H,CAO J,et al. Edge computing—an emerging computing model for the internet of everything era[J]. Journal of Computer Research and Development,2017,54 (5):907-924.

[5] WANG R,PEI X,ZHU J,et al. Multivariable time series forecasting using model fusion[J]. Information Sciences,2022,585:262-274.

[6] WANG R J,LAI J S,ZHANG Z Y,et al. Privacy-Preserving Federated Learning for Internet of Medical Things under Edge Computing[J/OL]. IEEE Journal of Biomedical and Health Informatics,2022. <https://ieeexplore.ieee.org/document/9729996>.

[7] KAIROUZ P,MCMAHAN H B,AVENT B,et al. Advances and open problems in federated learning[J]. Foundations and Trends® in Machine Learning,2021,14(1/2):1-210.

[8] YANG Q. Federated learning and artificial intelligence[J]. Software and Integrated Circuit,2019,5(11):961-965.

[9] KHAN L U,SAAD W,HAN Z,et al. Federated learning for internet of things:Recent advances,taxonomy, and open challenges [J]. IEEE Communications Surveys & Tutorials,2021,23(3):1759-1799.

[10] KONEČNÝ J,MCMAHAN H B,YU F X,et al. Federated learning:Strategies for improving communication efficiency[J]. arXiv:1610.05492,2016.

[11] ZHANG Y L,CHEN Y X,TIAN H,et al. Survey on Federated Learning Application on Scenarios of Edge Computing[J]. Journal of Chinese Computer Systems,2021,42(12):2645-2653.

[12] ZHAO Y,LI M,LAI L,et al. Federated learning with non-iid data[J]. arXiv:1806.00582,2018.

[13] MCMAHAN B,MOORE E,RAMAGE D,et al. Communication efficient learning of deep networks from decentralized data: Artificial intelligence and statistics[C]//PMLR. 2017.

[14] LI T,SAHU A K,ZAHEER M,et al. Federated optimization in heterogeneous networks[C]//Proceedings of Machine Learning and Systems. 2020:429-450.

[15] WANG J,LIU Q,LIANG H,et al. Tackling the objective inconsistency problem in heterogeneous federated optimization[J]. Advances in Neural Information Processing Systems,2020,33:7611-7623.

[16] KARIMIREDDY S P,KALE S,MOHRI M,et al. Scaffold:Stochastic controlled averaging for federated learning:International Conference on Machine Learning[C]//PMLR. 2020.

[17] PENG X,HUANG Z,ZHU Y,et al. Federated adversarial domain adaptation[J]. arXiv:1911.02054,2019.

[18] SERRA J,SURIS D,MIRON M,et al. Overcoming catastrophic forgetting with hard attention to the task:International Conference on Machine Learning[C]//PMLR. 2018.

[19] QU L,ZHOU Y,LIANG P P,et al. Rethinking Architecture Design for Tackling Data Heterogeneity in Federated Learning [J]. arXiv:2106.06047,2021.

[20] ZHANG M,SAPRA K,FIDLER S,et al. Personalized federated learning with first order model optimization[J]. arXiv:2012.08565,2020.

[21] LI L,DUAN M,LIU D,et al. FedSAE: A novel self-adaptive federated learning framework in heterogeneous systems:2021 International Joint Conference on Neural Networks (IJCNN) [C]//IEEE. 2021.

[22] REISIZADEH A,TAHERI H,MOKHTARI A,et al. Robust and communication-efficient collaborative learning [C] // Advances in Neural Information Processing Systems. 2019.



CHENG Fan, born in 1999, postgraduate, is a member of China Computer Federation. His main research interests include federated learning and edge computing.



WANG Rui-jin, born in 1980, Ph.D, associate professor, is a member of China Computer Federation. His main research interests include edge computing, blockchain, artificial intelligence and information security.