

中图法分类号: TP181 文献标识码: A 文章编号: 1006-8961(2024)12-3727-12

论文引用格式: Qi Y C, Huo Y L, Wang N and Hou Y. 2024. Joint dynamic correction algorithms for local and global drifts in federated learning. Journal of Image and Graphics, 29(12):3727-3738(戚银城, 霍亚琳, 王宁, 侯禹. 2024. 联邦学习中局部和全局偏移的联合动态校正算法. 中国图象图形学报, 29(12):3727-3738)[DOI:10.11834/jig.230891]

联邦学习中局部和全局偏移的联合动态校正算法

戚银城^{1,2*}, 霍亚琳¹, 王宁¹, 侯禹³

1. 华北电力大学电子与通信工程系, 保定 071003; 2. 华北电力大学河北省电力物联网技术重点实验室, 保定 071003;
3. 国网湖北综合能源服务有限公司武汉分公司, 武汉 430014

摘要: 目的 在联邦学习场景中, 由于各客户端数据分布的不一致, 会导致各客户端的局部目标之间偏差较大, 以及全局平均模型偏离全局最优, 影响模型训练的收敛速度和模型精度。针对非独立同分布数据导致的全局模型收敛缓慢以及模型准确率较低的问题, 提出一种联合动态校正的联邦学习算法(federated learning algorithm for joint dynamic correction, FedJDC), 分别从客户端和服务端进行优化。方法 为了降低局部模型更新偏移的影响, 定义累积偏移度来衡量各参与客户端的数据非独立同分布程度, 并在本地损失函数中引入动态约束项, 根据累积偏移度动态调整约束项大小, 可自动适应不同程度的非独立同分布数据, 减小局部模型的更新方向不一致性, 从而提高模型准确率及通信效率; 其次, 针对全局模型聚合偏移, 将参与客户端上传的累积偏移度作为全局模型聚合权重, 从而动态更新全局模型, 大幅减少通信轮数。结果 本文在3个真实数据集上的实验结果表明, 与4种不同的联邦学习算法相比, 在多种不同非独立同分布程度的情况下, FedJDC可以平均减少62.29%、20.90%、24.93%和20.47%的通信轮次, 平均提高5.48%、1.62%、2.10%和2.28%的模型准确率。结论 本文提出的联邦学习中局部和全局偏移的联合动态校正算法从局部模型更新和全局模型聚合两方面进行改进, 降低了通信轮次, 提高了准确率, 取得了良好的收敛效果。

关键词: 联邦学习(FL); 非独立同分布(non-IID); 损失函数; 模型聚合; 收敛性

Joint dynamic correction algorithms for local and global drifts in federated learning

Qi Yincheng^{1,2*}, Huo Yalin¹, Wang Ning¹, Hou Yu³

1. Department of Electronic and Communication Engineering, North China Electric Power University, Baoding 071003, China;
2. Hebei Key Laboratory of Power Internet of Things Technology, North China Electric Power University, Baoding 071003, China;
3. Wuhan Branch, State Grid Hubei Comprehensive Energy Service Co., Ltd., Wuhan 430014, China

Abstract: Objective Federated learning enables multiple parties to collaboratively train a machine learning model without communicating their local data. In practical applications, the data between nodes usually follow a non-independent identical distribution (non-IID). In the local update, each client model will be optimized toward its local optima (i. e., fitting its individual feature distribution) instead of the global optimal objective and raises a client update drift. Meanwhile, in global updates that aggregate these diverged local models, the server model is further distracted by the set of mismatching

收稿日期: 2023-12-27; 修回日期: 2024-03-06; 预印本日期: 2024-03-13

* 通信作者: 戚银城 qiyech@ncepu.edu.cn

基金项目: 河北省省级科技计划资助(SZX2020034)

Supported by: Hebei Provincial Science and Technology Plan(SZX2020034)

local optima, which subsequently leads to a global drift at the server model. To solve the problems of slow global convergence and increasing number of training communication rounds caused by non-IID data, this paper proposes a joint dynamic correction federated learning algorithm (FedJDC) that is optimized from the client and server. **Method** To reduce the influence of non-IID on federated learning, this paper carries out a joint optimization from the two aspects of local model update and global model update and proposes the FedJDC algorithm. This paper then uses the cosine similarity between the local and global update directions to measure the offset of each participating client. Afterward, given that each client has a different degree of non-IID, if the degree of the model offset is only determined by the cosine similarity calculated in this round, then the model update may become unstable. Therefore, the FedJDC algorithm defines the cumulative offset and introduces the attenuation coefficient ρ . In calculating the cumulative offset of the model, the current and historical cumulative offsets are taken into account. In addition, by changing ρ to reduce the proportion of the cumulative offset of the current round, the influence of the offset of the current round on the final result can be reduced. This paper also proposes a strategy for dynamically adjusting the constraint terms for local model update offset. Specifically, the constraint terms of the local loss function are dynamically adjusted according to the calculated cumulative offset of the local model, and the algorithm is automatically adapted to various non-IID settings without a careful selection of hyperparameters, thus improving the flexibility of the algorithm. To dynamically change the weight of global model aggregation in each round and effectively improve the convergence speed and model accuracy, this paper also designs a dynamic weighted aggregation strategy that takes the accumulated offset uploaded by all clients as the weight of global model aggregation in each round of communication. **Result** The proposed method is tested on three dataset using different deep learning models. LeNet5, the VGG16 network model, and the ResNet18 network model are used for training in the MNIST, FMNIST, and CIFAR10 datasets, respectively. Four experiments are designed to prove the effectiveness of the proposed algorithm. To verify the accuracy of FedJDC at different degrees of non-IID, the hyperparameter β of the Dirichlet distribution is varied, and the performance of different algorithms is compared. Experimental results show that FedJDC can improve the model accuracy by 5.48%, 1.62%, 2.10%, and 2.28% on average compared with FedAvg, FedProx, FedAdp, and FedLAW, respectively. To evaluate the communication efficiency of FedJDC, the number of communication rounds is counted as FedJDC reaches a target accuracy, and this number is compared with that obtained by other algorithms. Experimental results show that under different degrees of non-IID, FedJDC can reduce communication rounds by 62.29%, 20.90%, 24.93%, and 20.47% on average compared with FedAvg, FedProx, FedAdp, and FedLAW, respectively. This paper also investigates the effect of the number of local epochs on the accuracy of the final model. Experimental results show that FedJDC outperforms the other four methods under different epochs in terms of final model accuracy. FedJDC also demonstrates better robustness against the larger offset caused by more local update epochs. Ablation experiments also show that each optimization method performs well on all datasets, and FedJDC combines these two strategies to achieve the global optimal performance. **Conclusion** This paper optimizes the local and global model offsets from two aspects and proposes a joint dynamic correction algorithm for these offsets in federated learning. The cumulative offset is defined, and the attenuation coefficient is introduced into the calculation of the cumulative offset. Considering the historical and current offset information, the size of the cumulative offset is dynamically adjusted to ensure the stability of the training parameter update. The dynamic constraint strategy takes the cumulative offset calculated by the client in each round as the constraint parameter of the client model. The dynamic weighted aggregation strategy changes the weight of each local model during the global model aggregation based on the cumulative offset of each participating client so as to dynamically update the global model in each round. The combination of the two optimization strategies has achieved good results, effectively alleviated the performance degradation of the federated learning model caused by non-IID data, and provided a good foundation for the further implementation of federated learning in this field.

Key words: federated learning (FL); non-independent identical distribution (non-IID); loss function; model aggregation; convergence

0 引言

众所周知,人工智能(artificial intelligence, AI)的力量来自于大数据,随着全球数据量的激增,AI发展迅速,而现实情况是许多数据分散在各地各行业各领域,部分领域之间的数据很难安全共享,这也导致了目前数据孤岛的问题。近年来,由于隐私泄露事件频发,人们对个人隐私的保护意识越来越强,许多国家也出台了一系列保护用户数据隐私的法规,如何在保护数据隐私前提下有效利用分布在各地的数据进行机器学习是一项挑战。

为维护隐私敏感数据并促进分布式节点之间的协作机器学习,谷歌(McMahan等,2017)提出了联邦学习(federated learning, FL)的概念,通过将本地训练的模型更新传输到中央服务器进行模型聚合,无需发送本地原始数据,有效地保护了用户数据隐私。

联邦学习吸引了不同应用领域的研究兴趣,如医学成像(Kaissis等,2020;Kumar等,2021)、视频分析系统(Li等,2020a)、目标检测(Liu等,2020)、地标分类(Hsu等,2020)等。但联邦学习在实际场景中落地还面临着诸多问题,其中最具挑战性的是不同节点数据分布的异质性(Li等,2020b)。由于参与联邦学习的客户端各自采集数据,且由于各设备所处环境及采集偏好不同,其数据往往分布差异较大且各自独立,因此数据异构在统计学上表现为数据是非独立同分布(non-independent identical distribution, non-IID)的,这限制了联邦学习的模型性能和效率(Li等,2022)。non-IID数据会使得各节点在更新其局部模型时,其局部目标之间差距较大,使得局部模型将朝着各自局部最优进行优化,这会在客户端更新中引起偏移(Zhao等,2022)。同时,各局部目标与全局目标之间也有偏差,局部模型没有朝着全局最优进行优化,并且服务器端在聚合这些分散的局部模型时,全局模型进一步被各客户端的局部最优分散,这将导致全局模型的偏移。模型偏移影响模型训练的收敛速度,甚至可能导致模型无法收敛,给训练过程带来了不稳定性。

为了解决non-IID数据引起的FL模型性能下降问题,现有的工作主要可以分为两组,分别对应于局部偏移和全局偏移。

1)解决本地客户端训练中的局部偏移问题。本

地优化方面主要强调修正局部模型更新方向,Li等人(2020a)为限制因数据异质性导致的用户模型偏离,提出了一种FedProx算法,在局部训练的目标函数中引入了一个近端项,为用户本地损失函数添加了正则项,其本质是限制用户本地模型和全局模型之间的差异。Karimireddy等人(2020)通过在本地模型的更新公式中添加一个修正项对模型的更新方向进行修正,使其朝着真正的最优解方向移动,并从理论上证明了所提随机控制平均方法(stochastic controlled averaging algorithm, SCAFFOLD)具有更高的收敛速率。然而,SCAFFOLD方法仅考虑降低通信代价,模型精度无法得到很好的保证。Li等人(2021a)提出在本地保留批规范化层,以规范化本地数据分布,从而帮助客户端获得相似的特征分布。Li等人(2021b)在客户端使用潜在特征表示的对比学习来增强本地和全局模型之间的一致性。

2)解决服务器端聚合的全局目标偏移问题。对于全局服务器端的优化,调整聚合权重是一种典型的策略,McMahan等人(2017)提出了联邦平均(federated averaging, FedAvg)算法,客户端利用自身数据在本地训练,并将训练后得到的模型参数上传至服务器,服务器根据客户端训练数据的数量对模型参数进行加权平均,再将模型下发给各客户端。但由于依赖non-IID数据的大量局部更新会使得模型朝向局部目标的最优,而不是全局目标的最优,局部模型和全局模型之间的不一致性随着局部训练的积累而积累,导致训练收敛前会有更多的通信回合,使得FedAvg的收敛速度降低。Huang等人(2020)设计了一个名为LoAdaBoost的FedAvg变体算法,通过比较其交叉熵损失与上一轮的中值损失来定义每个客户端的拟合程度,从而动态调整当前迭代中每个客户端本地训练轮次。Wang等人(2020a)利用贝叶斯非参数方法分层匹配和平均权重。Hsu等人(2019)在服务器上更新全局模型时应用动量。Wang等人(2020b)在平均之前对局部更新进行了标准化。Wu和Wang(2021)提出了FedAdp(federated learning with adaptive weighting)算法,基于局部梯度和全局梯度夹角量化节点贡献,在每轮聚合中根据节点的贡献而不是局部数据集的大小来自适应地分配不同权重,以减少non-IID数据下FL训练的通信轮数。Li等人(2023)提出一种有效的具有可学习聚合权重的联邦学习方法FedLAW(federated learning with learn-

able aggregation weights), 该算法可以在不同的数据集和模型上大幅度提高全局模型的泛化能力。

这些方法通过优化局部或全局更新来部分地解决 non-IID 数据所带来的问题。但本地和全局的偏移本质上是互相影响互相耦合的, 如何作为一个整体综合考虑以解决数据的非独立同分布问题仍是一大难题。

本文为了克服传统联邦学习算法在 non-IID 场景下存在的模型偏移问题, 从局部模型更新和全局模型更新两方面进行优化, 提出联合动态校正联邦学习算法 (federated learning algorithm for joint dynamic correction, FedJDC)。为了更直观地衡量模型更新方向的偏移程度, 本文定义了累积偏移度, 并将其作为局部模型约束项以及全局模型聚合时的权重, 从而进行局部和全局联合动态调整, 加快模型收敛速度, 提高模型准确率。

本文的主要贡献如下: 1) 通过观察联邦学习模型在 non-IID 数据场景下的更新方向, 可以发现数据 non-IID 程度越强, 各局部模型之间以及局部模型与全局模型之间的更新方向偏差越大。因此, 本文采用局部模型更新方向与全局模型更新方向之间的余弦相似度来反映每个局部模型与全局模型更新方向之间的偏移程度, 并且在累积偏移度的计算中引入衰减系数, 同时考虑了历史和当前的偏移度信息, 从而动态调整累积偏移度的大小, 保证了训练参数更新的稳定性。2) 在 FedJDC 算法中, 针对局部模型更新方向偏移问题, 提出一种动态约束策略, 在本地损失函数中引入约束项, 并根据计算出的累积偏移度动态调整约束项大小。同时, 针对全局模型更新偏移问题, 设计了一种动态加权聚合策略, 将参与客户端上传的累积偏移度作为其在每一轮通信中参与全局模型聚合时的权重, 从而动态改变每一轮全局模型聚合时的权重, 有效提高收敛速度和模型准确率, 使算法更加灵活。3) 通过在 3 个真实数据集上的大量实验, 对所提出的 FedJDC 算法的性能进行了评估。实验结果表明, 与其他改进算法相比, 使用 FedJDC 进行联邦学习模型训练可以大大减少通信轮数并提高模型准确率。

1 研究方法

为降低 non-IID 数据对联邦学习模型收敛速度

的影响, 本算法从本地客户端局部模型更新和服务端全局模型聚合两方面对 FedAvg 算法进行改进, 提出了结合动态约束策略和动态加权聚合策略的联合动态校正的联邦学习算法。

1.1 联邦学习架构

边缘设备利用本地数据进行联邦学习的网络架构如图 1 所示, 其中包括 M 个边缘设备作为客户端和一个用来训练全局模型的服务器, 每个设备所拥有的数据样本数为 $D_i, i \in [1, M]$ 。设每轮通信随机选择 N 个客户端参与训练, 则联邦学习的优化目标为

$$\min_w F(w) = \sum_{i=1}^N \lambda_i F_i(w) \quad (1)$$

式中, $F_i(w)$ 为局部目标函数, $F(w)$ 为全局目标函数, w 为全局模型, λ_i 为聚合权重, 满足 $\sum_{i=1}^N \lambda_i = 1$ 。在联邦平均算法 (FedAvg) 中, 每个客户端在本地执行多轮随机梯度下降法 (stochastic gradient descent, SGD) 更新, 服务器再进行全局聚合, 每一轮联邦学习分为两个阶段:

1) 本地训练 (客户端)。每个参与客户端执行随机梯度下降 (SGD) 训练, 以优化其局部目标 $F_i(w)$, 具体为

$$w_i(t) = w(t-1) - \eta \nabla F_i(w(t-1)) \quad (2)$$

式中, η 为学习率, $\nabla F_i(\cdot)$ 为客户端 i 的梯度。经过 E 个 epoch 之后, 客户端将本地模型更新 $\Delta_i(t) = w_i(t) - w_i(t-1)$ 发送给服务器。

2) 全局聚合 (服务器端)。中央服务器将各个客户端上传的模型参数聚合并更新相应的全局模型, 即

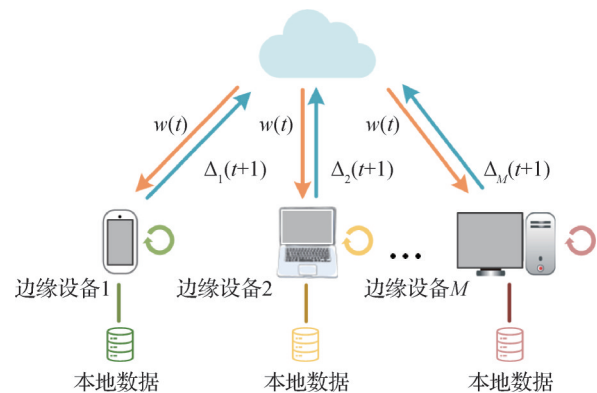


图1 联邦学习架构

Fig. 1 Framework of a federated learning

$$\Delta(t) = \sum_{i=1}^N \lambda_i \Delta_i(t) \quad (3)$$

$$w(t) = w(t-1) + \Delta(t) \quad (4)$$

式中, $\lambda_i = D_i / \sum_{i=1}^N D_i$, 然后将全局模型参数分配给选定的客户端, 进行下一步局部训练和全局聚合。

1.2 non-IID 数据造成的模型偏移

在数据 non-IID 场景下, 由于每个客户端数据集之间的分布差异很大, 同时与全局分布差异也很大, 因此, 各局部最优目标与全局最优目标不一致, 这就会导致模型训练效率降低, 甚至在实践过程中间停止训练。图2演示了 IID 数据和 non-IID 数据设置下的 FedAvg 更新。在 IID 设置下, 全局最优值 w^* 接近局部最优值 w_1^* 和 w_2^* 。因此, 平均模型 $w(t+1)$ 也接近于全局最优。然而, 在 non-IID 场景下, 由于不同客户端所拥有的数据分布不同, 各客户端均朝本地子问题最优解进行更新, 造成各局部模型更新方向的偏移。并且 w^* 与 w_1^* 、 w_2^* 相距较远, 所以平均模型 $w(t+1)$ 与全局最优 w^* 之间的距离也较远。各客户端更新方向的偏移以及平均模型 $w(t+1)$ 与全局最优 w^* 之间的差异会导致模型的收敛速度降低。

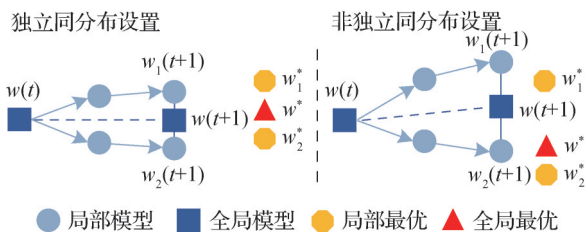


图2 non-IID 设置下的模型偏移

Fig. 2 Illustration of model offset under non-IID settings

1.3 局部模型的动态约束策略

为了取得更好的收敛稳定性, FedProx 提出在局部模型训练损失函数中添加一个距离约束项, 限制局部更新与全局模型之间的距离, 缓解模型偏移问题。约束权重 μ 用于控制局部模型与全局模型之间的差异, 所有客户端采用相同的权重 μ 。FedProx 方法的收敛性能相比于 FedAvg, 在大多数情况下并没有明显优势, 并且需要对约束权重 μ 进行精细的调试。袁强强等人(2010)将数值计算领域的 U 曲线方法引入到超分辨率重建领域, 用来动态调整模型中的最优正则化参数。本算法提出一种动态调整约束项的策略, 根据计算出的局部模型累积偏移度动态

调整本地损失函数的约束项, 自动适应各种 non-IID 设置, 而无需精心选择超参数, 可有效提升联邦学习模型在 non-IID 数据分布情况下的收敛速度以及模型准确率。

1.3.1 累积偏移度指标定义

在 non-IID 数据场景下, 各客户端均向本地子问题最优解进行更新, 造成各局部模型更新方向的偏移, 并且聚合后的全局模型与全局最优之间的距离也较远。余弦相似度(Xia等, 2015)是用两个向量之间夹角的余弦值来衡量两个个体之间差异的大小, 即相似度的大小。因此, 可采用余弦相似度作为模型偏移度的衡量指标, 具体计算为

$$\cos \theta_i = \frac{\langle w_i - w(t), w(t) - w(t-1) \rangle}{\|w_i - w(t)\| \times \|w(t) - w(t-1)\|} \quad (5)$$

式中, $\langle \cdot \rangle$ 代表内积运算。

每个客户端所拥有的 non-IID 程度不同, 若模型的偏移程度仅由本轮计算出的余弦相似度决定, 则可能会为模型更新带来不稳定性。因此, 本算法引入了衰减系数 ρ , 在模型累积偏移度计算阶段不仅考虑当前的累积偏移度, 也考虑了历史累积偏移度, 并且改变 ρ 值减少本轮累积偏移度所占比例, 从而降低该轮次偏差对最后结果的影响, 定义累积偏移度 $\varphi_i(t)$ 指标为

$$\varphi_i(t) = \rho \times \varphi_i(t-1) + (1 - \rho) \times \cos \theta_i \quad (6)$$

式中, $\varphi_i(t)$ 和 $\varphi_i(t-1)$ 分别代表客户端 i 在当前和上一轮的累积偏移度, ρ 表示衰减系数, $\cos \theta_i$ 表示局部模型更新方向和全局模型更新方向之间的余弦相似度。由式(6)可知, 每一次累积偏移度 $\varphi_i(t)$ 计算都会考虑前面的累积值, ρ 值越大, 本轮偏移度所占权重越小, 并且通过对衰减系数 ρ 进行调优, 本文得到最佳 ρ 值为 0.7。因此, 若本轮累积偏移度和历史累积偏移度相差较多, 引入衰减系数可以降低偏差对本轮最后结果的影响, 从而保证整个优化方向的稳定性。

1.3.2 动态约束策略目标函数设计

在 non-IID 场景下, 每个客户端所拥有的数据分布不同, 每一轮局部模型的更新方向偏移程度不同, 在对局部训练添加距离约束时, 不同偏移程度的模型更新方向需要不同的正则化约束系数。当客户端本地模型与全局模型更新方向较为一致的时候, 可以适当减小约束, 加快收敛速度; 反之, 则可以增大

约束,从而有效地限制模型偏移。针对以上问题,提出一种联邦动态约束策略来改善联邦学习模型的性能。在原有本地损失函数的基础上加入了约束项,巧妙地应用了全局模型和本地模型的动态关系缓解 non-IID 问题,定义目标函数为

$$\arg \min_w h_i(w; w(t)) = F_i(w) + \sigma \times (1 - \varphi_i(t)) \quad (7)$$

式中, h_i 表示优化目标函数,约束权重 σ 用于控制局部模型与全局模型之间的差异, $\varphi_i(t)$ 为累积偏移度。

动态约束策略在局部训练阶段,为每个客户端增加了一个方向惩罚,并且根据计算出的累积偏移度可以动态计算出约束项参数,减少调参工作量和一些冗余计算,加快收敛速度,从而得到更快更高效的联邦学习模型。

1.4 全局模型的动态加权聚合策略

由 1.2 节所描述的问题可以分析出,减少全局模型融合时的方向偏移将是减轻 non-IID 数据影响的一个关键问题。由于模型参数的更新本质上是基于损失函数梯度的更新,所以联邦模型的聚合本质上也是各个参与客户端梯度更新方向的融合,因此,当某个客户端数据的 non-IID 程度越大时,其产生的梯度更新方向偏差也越大,该客户端对全局模型所贡献的更新梯度就应被更大程度地限制,从而使得模型偏移减小,保障 non-IID 数据场景下联邦学习模型的性能。

由式(5)可知,余弦相似度数较大时,说明局部更新方向与全局梯更新方向相似,有利于全局聚合。相反,当余弦相似度数较小时,例如等于或小于 0 时,则局部更新方向与全局梯更新方向相反,从而对全局聚合产生负影响。因此,本文提出动态加权聚合策略,利用客户端计算的累积偏移度 $\varphi_i(t)$ 来反映每个参与客户端的贡献。

在对每个客户端的累积偏移度进行归一化后,最终得到全局模型聚合参与客户端的权值为

$$\lambda_i(t) = \begin{cases} \frac{\varphi_i(t)}{\sum_{i'=1}^N \varphi_{i'}(t)} & D_m = D_n \quad \forall m, n \in [1, N] \\ \frac{D_i \varphi_i(t)}{\sum_{i'=1}^N D_{i'} \varphi_{i'}(t)} & D_m \neq D_n \quad \exists m, n \in [1, N] \end{cases} \quad (8)$$

如果所有参与客户端都具有相同数量的数据样本,则将通过累积偏移度量化的贡献来分配权重。反之,将根据累积偏移度和客户端所拥有的数据集

样本数量分配权重。在 FedAvg 中,聚合的权重仅与局部数据集的数量成正比,而动态加权聚合策略在为模型聚合分配权重的同时考虑了客户端数据集样本数量和客户端模型更新方向的累积偏移度,可以有效缓解 non-IID 数据给全局模型更新带来的偏移问题。

结合以上两种策略,从而得到联合动态校正算法,通过量化局部更新方向和全局更新方向之间的余弦相似度来衡量客户端模型相对于全局模型的偏移程度。并且定义了累积偏移度指标,局部模型根据计算出的累积偏移度动态限制更新方向,缓解局部模型偏移;全局模型更新的权重也可以根据客户端上传的累积偏移度动态分配,而不是简单根据数据集的大小进行加权平均,具体算法流程如算法 1 所示。

算法 1 联合动态校正的联邦学习算法 FedJDC。

输入: 批量处理大小 B , 通信总轮数 T , 学习率 η , 总客户端数 M 。

1) 服务器随机从总客户端数 M 中选择 N 个客户端(每个设备的选择概率为 p_k); 服务器初始化模型参数 $w(0)$ 和全局模型更新 $\Delta(0)$ 到所有被选中客户端;

2) for $t = 1, 2, \dots, T-1$ do;

3) for $i \in [1, N]$ do;

4) $\Delta_i(t) \leftarrow \text{Local Update}(i, w(t-1))$;

5) $w(t) \leftarrow \text{Global Update}(\Delta_i(t), \varphi_i(t))$;

局部客户端更新(local update):

输入: 客户端 i , 全局模型参数 $w(t)$, 全局更新参数;

6) 每个客户端 $i \in [1, N]$ 通过最小化式(7)来更新本地模型 $w_i(t)$;

7) 计算模型参数差 $\Delta_i(t) = w_i(t) - w(t-1)$;

8) 输出 $\Delta_i(t)$, $\varphi_i(t)$;

服务器端更新(global update):

输入: 累积偏移度 $\varphi_i(t)$, 归一化后的累积偏移度 $\lambda_i(t)$, 局部更新参数 $\Delta_i(t)$;

9) 聚合全局模型更新参数 $\Delta(t) = \sum_{i=1}^N \lambda_i(t) \Delta_i(t)$;

10) 更新全局模型 $w(t) \leftarrow w(t-1) + \Delta(t)$;

11) 输出 $w(t)$, $\Delta(t)$ 。

2 实验结果与分析

为了评估本文提出的 FedJDC 算法性能,实验在

3个真实数据集上进行图像分类实验,并且在不同程度的 non-IID 数据下与基线算法 FedAvg、当前具有代表性的客户端正则化方法 FedProx、自适应加权算法 FedAdp 以及可学习聚合权重方法 FedLAW 进行对比。

2.1 实验设置

在配置有 GeForce RTX 2080 Ti 的服务器上进行模拟仿真,并使用 PyTorch 框架实现 FedJDC 算法。本文分别在 MNIST(modified National Institute of Standards and Technology)(Lecun 等,1998),FMNIST(fashion-modified National Institute of Standards and Technology)(Cohen 等,2017)和 CIFAR10(Canadian Institute for Advanced Research 10)(Krizhevsky,2009)3个数据集上进行了实验,数据集的统计数据汇总见表 1。对于 MNIST 数据集,使用 LeNet5(LeCun’s Network 5)进行训练;对于 FMNIST 数据集,使用 VGG16(Visual Geometry Group Network 16-layer)网络模型进行训练;对于 CIFAR10 数据集,使用 ResNet18(residual network 18-layer)网络模型进行训练。总客户端数量 $M=100$,每轮参与训练的客户端数量 $N=10$,每个客户端的数据样本数量 $D_i=$

600,批量处理大小 $B=32, E=5$,初始学习率 $\eta=0.01$,服务器端动量系数为 0.5。参考 Yurochkin 等人(2019)提出的 non-IID 数据划分方式,本文算法同样使用狄利克雷分布来实现每个客户端 non-IID 数据的划分。狄利克雷分布通常用做贝叶斯统计中的先验分布,是模拟现实世界数据分布的合适选择。具体来说,本算法使用 $p_k \sim \text{Dir}_N(\beta)$ 进行采样,将第 k 类的 $p_{k,j}$ 比例的数据分配给客户端 j ,其中 $\text{Dir}(\cdot)$ 表示狄利克雷分布, β 为浓度参数。通过改变参数 β 可以灵活地改变数据非独立同分布的程度。图 3 展示了 10 个客户端在两个不同 β 取值下进行数据随机采样的分布情况,可以看出 β 值越小时,客户端拥有的类别数量越少且类别间的数量越不平衡。

表 1 实验数据集统计
Table 1 Statistics of experimental datasets

数据集	训练样本/个	测试样本/个	特征/像素	类别/个
MNIST	60 000	10 000	784	10
FMNIST	60 000	10 000	784	10
CIFAR10	50 000	10 000	1 024	10

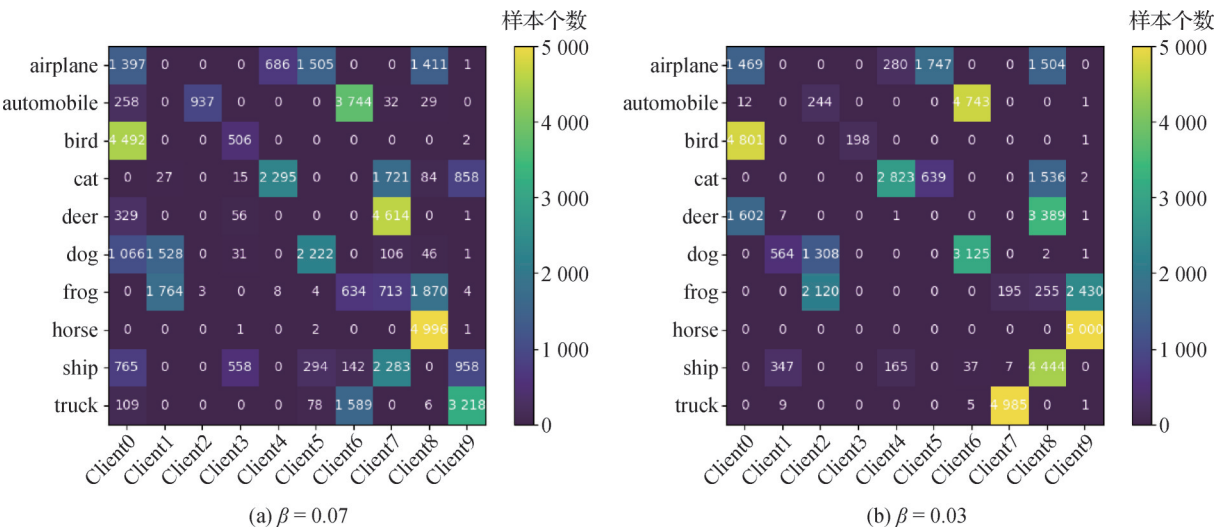


图 3 CIFAR10数据集上基于狄利克雷分布划分示例

Fig. 3 Examples of partitioning based on the Dirichlet distribution on the CIFAR10 dataset ((a) $\beta = 0.07$; (b) $\beta = 0.03$)

2.2 不同非独立同分布设置下的算法性能

为了验证 FedJDC 的有效性,本实验通过改变 Dirichlet 分布的浓度参数 β 来研究不同 non-IID 程度下各算法的性能。 β 越小,数据分配越不平衡,non-IID 程度越大,对模型收敛影响越大。不同 non-IID 程度下的测试准确率如图 4 所示。

由图 4 可以看出,随着 β 的变小,FedAvg 在 MNIST 数据集上的收敛速度降低,测试精度也随之降低。本实验选用了 $\beta \in \{0.07, 0.05, 0.03\}$ 对各算法进行测试,并对比在 3 个真实数据集上的平均测试准确率。对于 FedJDC 中的超参数 σ ,从 $\{0.002, 0.02, 0.2, 2\}$ 中调优,并报告最佳结果。

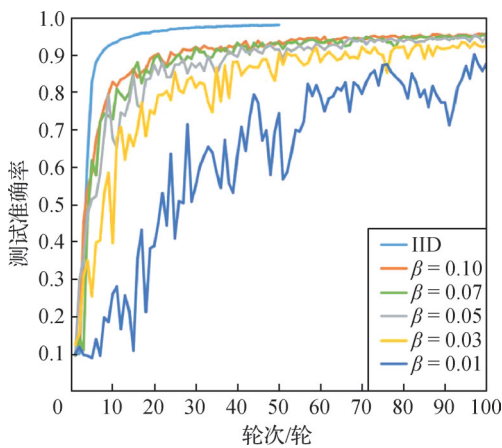


图4 不同 non-IID 程度下的测试准确率
Fig. 4 Test accuracy at different levels of non-IID

MNIST, FMNIST 和 CIFAR10 的最佳 σ 值均为 0.02。对于 FedProx 中的超参数,按照其原文所提出的 μ 取值范围 $\{1, 0.1, 0.01, 0.001\}$ 来调整不同数据集对应的最优参数。通过实验得出 MNIST、FMNIST 和 CIFAR10 这 3 个数据集在 FedProx 算法中的最佳 μ 值

分别为 1、0.01 和 0.001。
表 2 展示了所有方法在默认设置下的测试准确率。它们的模型收敛精度曲线分别如图 5 所示,图中显示了训练过程中每一轮的准确性,随着客户端数据 non-IID 程度增加,算法之间的效果差异开始体现。首先可以很明显地看出,在异质性数据场景下, FedAvg 算法性能下降明显。在 FedJDC 算法的效果方面,可以观察到代表 FedJDC 算法的模型准确率曲线较 FedAvg 以及其他对照方法不仅可以获得最高的准确率,而且具有更快收敛到最高准确率点的速率。改进后的算法在 3 个数据集上与基线算法 FedAvg (McMahan 等, 2017) 相比测试精度平均高出 5.48%, 与 FedLAW (Li 等, 2023) 相比测试精度平均高出 2.28%, 且可达到更快更稳的收敛效果。由此可知,本文提出的 FedJDC 可以有效缓解 non-IID 带来的局部模型之间以及局部与全局更新方向之间偏差较大的问题,在实验中能够减少模型精度的振荡,使其更快地朝着更加稳定的全局最优解方向收敛。

表 2 不同 non-IID 程度下的准确率比较
Table 2 Comparison of accuracy under different levels of non-IID

方法	/%								
	MNIST数据集			FMNIST数据集			CIFAR10数据集		
	$\beta = 0.07$	$\beta = 0.05$	$\beta = 0.03$	$\beta = 0.07$	$\beta = 0.05$	$\beta = 0.03$	$\beta = 0.07$	$\beta = 0.05$	$\beta = 0.03$
FedAvg 基线 (McMahan 等, 2017)	94.12	93.31	92.37	88.11	85.81	80.12	61.73	57.60	55.50
FedProx (Li 等, 2020a)	96.43	96.22	94.95	88.07	87.52	86.95	67.01	63.46	62.83
FedAdp (Wu 和 Wang, 2021)	96.63	96.10	94.93	87.76	86.90	85.81	66.36	66.16	58.46
FedLAW (Li 等, 2023)	96.73	96.49	95.63	88.86	87.93	87.25	62.81	61.03	60.66
FedJDC (本文)	96.89	96.86	96.59	89.85	89.68	88.25	68.24	68.06	63.56

注:加粗字体表示各列最优结果。

2.3 通信效率

为了进一步验证 FedJDC 算法的有效性,与表 2 相同的设置,首先执行 100 轮的其他主流改进算法。然后,与优化后的方法训练相同性能模型所需的通信轮数进行对比。
由表 3—表 5 可知,在 3 种数据集的不同 non-IID 程度设置下,本文所提出的 FedJDC 算法与 FedAvg 相比,训练所需要的通信轮数可减少一半以上,且与其他 3 种改进方法相比平均可以减少 20.90%、24.93% 和 20.47% 的通信轮次,通信效率大幅提高。同时,数据 non-IID 程度越高, FedJDC 的优势越明

显。通过对实验结果分析可知, FedJDC 可以通过从局部和全局两方面对模型偏移进行校正,从而减轻其模型更新方向的偏移程度,提高非独立同分布场景下联邦学习模型的通信效率。
2.4 不同局部更新 epoch 的影响
不太频繁的通信会进一步增强 non-IID 数据导致的漂移,从而使得聚合后的全局模型表现更差。本文为了研究不同局部更新 epoch 下 FedJDC 的有效性,选用了 CIFAR10 数据集以及较为极端的 non-IID 数据分布 $\beta = 0.03$ 进行实验,若在此设置下 FedJDC 仍表现良好,则在其他较小数据集且 non-IID 程度较

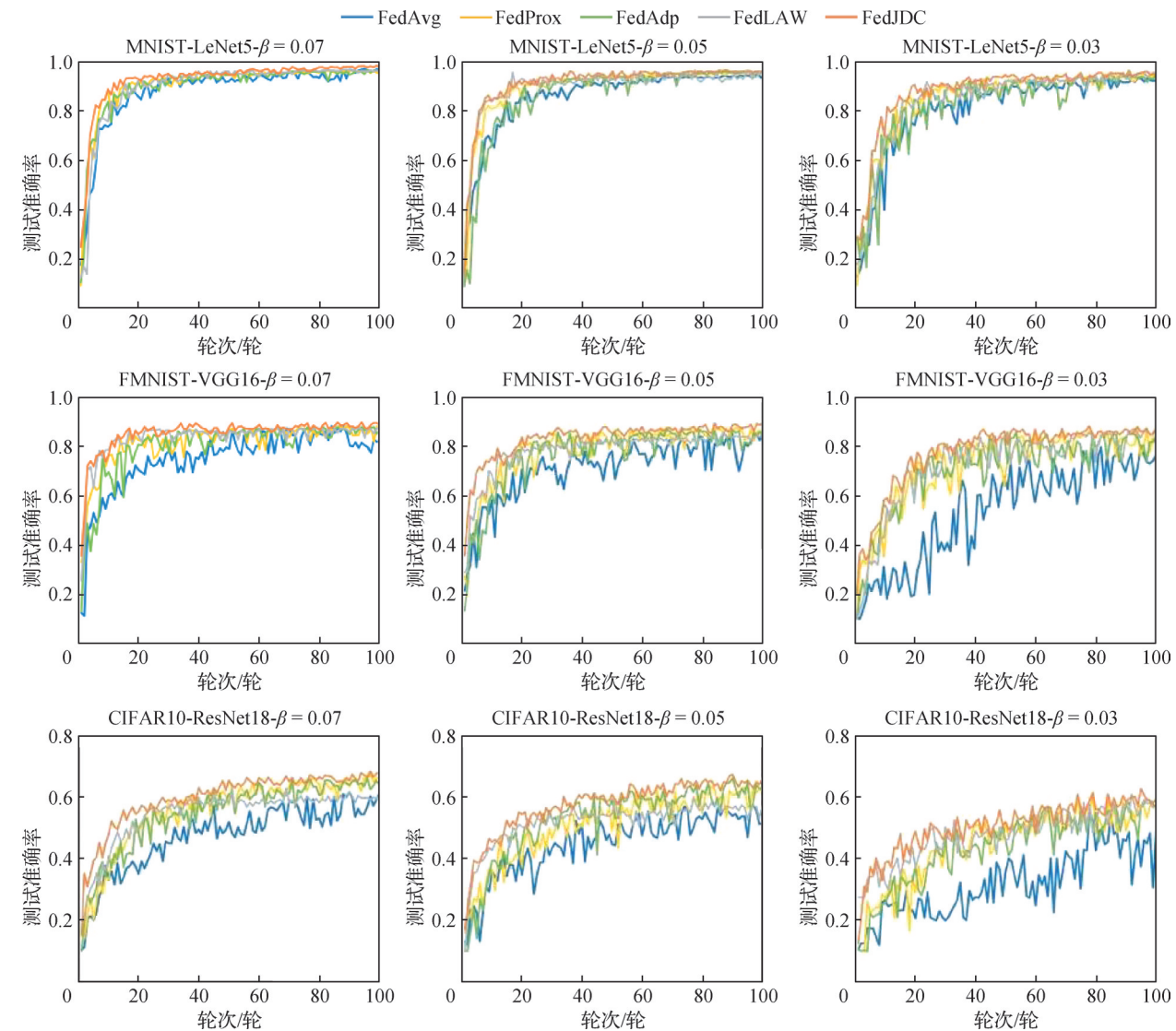


图5 不同算法的模型收敛性及准确率对比

Fig. 5 Comparison of model convergence and accuracy of different algorithms

表 3 在 MNIST 数据集训练中达到目标精度 90% 的通信轮次

Table 3 Communication rounds that achieve 90% of the target accuracy in training on the MNIST dataset

方法	通信轮次/轮		
	$\beta = 0.07$	$\beta = 0.05$	$\beta = 0.03$
FedAvg 基线 (McMahan 等, 2017)	24	38	51
FedProx (Li 等, 2020a)	13	18	29
FedAdp (Wu 和 Wang, 2021)	14	16	27
FedLAW (Li 等, 2023)	15	17	24
FedJDC (本文)	11	13	18

注:加粗字体表示各列最优结果。

轻的场景下应会得到更加良好的实验结果。

实验结果如图 6 所示,当局部更新 epoch 数 $E = 1$ 时,局部更新非常小。因此,在通信回合数相同的情

况下,训练速度较慢,准确率相对较低。所有的方法都有相近的精度,但优化后的算法仍然是表现最好的。当 E 过大时,所有方法的精度都会下降,这是由

表 4 在 FMNIST 数据集训练中达到目标精度 80% 的通信轮次
Table 4 Communication rounds that achieve 80% of the target accuracy in training on the FMNIST dataset

方法	通信轮次/轮		
	$\beta = 0.07$	$\beta = 0.05$	$\beta = 0.03$
FedAvg 基线 (McMahan 等, 2017)	26	38	52
FedProx (Li 等, 2020a)	15	18	29
FedAdp (Wu 和 Wang, 2021)	17	21	38
FedLAW (Li 等, 2023)	16	23	27
FedJDC (本文)	10	16	26

注:加粗字体表示各列最优结果。

表 5 在 CIFAR10 数据集训练中达到目标精度 40% 的通信轮次
Table 5 Communication rounds that achieve 40% of the target accuracy in training on the CIFAR10 dataset

方法	通信轮次/轮		
	$\beta = 0.07$	$\beta = 0.05$	$\beta = 0.03$
FedAvg 基线 (McMahan 等, 2017)	44	57	95
FedProx (Li 等, 2020a)	17	20	38
FedAdp (Wu 和 Wang, 2021)	16	24	34
FedLAW (Li 等, 2023)	14	26	30
FedJDC (本文)	13	19	29

注:加粗字体表示各列最优结果。

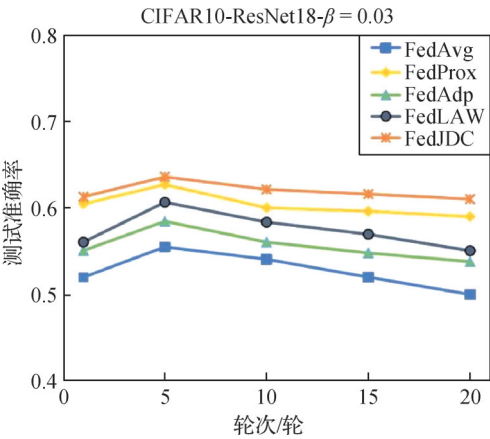


图 6 在不同局部训练 epoch 下的比较
Fig. 6 Comparison under different local training epochs

于不频繁的通信会进一步增强 non-IID 特征的漂移,最终得到更差的全局模型。然而,FedJDC 显然优于其他方法,并且对更多局部更新 E 带来的更大漂移显示出鲁棒性。

2.5 消融实验

FedJDC 算法从两方面进行了优化,为了验证两个优化策略分别对算法的影响,本文选择进行消融

实验。分别对两个优化策略以及结合两个策略后的 FedJDC 算法进行实验,采用 $\beta = 0.03$ 的 non-IID 数据,表 6 展示了在 3 种数据集上的消融实验结果。

通过实验结果可以看出,两个单一优化策略的模型精度均在 FedJDC 与基线算法 FedAvg 之间,都能带来模型精度的提高,并且在 FMNIST 数据集上的表现更为明显,与基线算法 FedAvg 相比,精度提升分别达到 7.92% 和 7.79%;而结合两个优化后的 FedJDC 算法精度提升可达 8.13%。这些结果充分表明,单独使用局部动态约束策略或全局动态加权聚合策略都有助于解决 non-IID 问题,而且结合两种优化的 FedJDC 算法相比于这两种方法单独使用的情况,可以达到最好的效果。

3 结 论

联邦学习已经成为解决数据孤岛问题的一种很有前景的方法,而 non-IID 数据是联邦学习的一个巨大挑战。为了提高联邦学习模型在 non-IID 数据场

表 6 两个优化策略分别与其他算法相比的测试准确率
Table 6 Test accuracy of the two optimization strategies compared to other algorithms respectively

方法	测试准确率/轮		
	MNIST	FMNIST	CIFAR10
FedAvg 基线(McMahan 等, 2017)	92.37	80.12	55.50
FedProx(Li 等, 2020a)	94.95	86.95	62.83
FedAdp(Wu 和 Wang, 2021)	94.93	85.81	58.46
FedLAW(Li 等, 2023)	95.63	87.25	60.66
局部动态约束策略	96.24	88.04	62.89
全局动态加权聚合策略	96.50	87.91	63.01
FedJDC(本文)	96.59	88.25	63.56

注:加粗字体表示各列最优结果。

景下的通信效率以及模型精度,本文针对局部和全局模型偏移问题提出了一种结合动态约束策略和动态加权聚合策略的联合动态校正联邦学习算法。本算法定义了累积偏移度,以定量表示模型更新方向的偏移程度,并将每一轮客户端计算出的累积偏移度作为局部模型约束项参数以及全局模型更新聚合时的权重,从而联合局部和全局动态调整参数大小。

通过在 3 个数据集上进行各种定量对比实验,所提出的联合动态校正的联邦学习算法与基线算法 FedAvg 以及其他联邦学习改进算法相比,可以根据各客户端的非独立同分布程度动态自适应改变参数,有效加快联邦模型的收敛速度,提高联邦模型的精度。在未来的工作中,可以考虑优化客户端选择方案,动态地改变每个客户端被选择参与训练的概率,对于数据 non-IID 程度强以及历史训练效果不理想的节点,增加其被选择的概率,使得模型尽可能快地学习全局数据的知识,从而加快联邦模型收敛,提升效率。

参考文献(References)

Cohen G, Afshar S, Tapson J and van Schaik A. 2017. EMNIST: extending MNIST to handwritten letters//Proceedings of 2017 International Joint Conference on Neural Networks. Anchorage, USA: IEEE: 2921-2926 [DOI: 10.1109/IJCNN.2017.7966217]
Hsu T M H, Qi H and Brown M. 2019. Measuring the effects of non-identical data distribution for federated visual classification [EB/OL]. [2023-12-01]. <https://arxiv.org/pdf/1909.06335.pdf>
Hsu T M H, Qi H and Brown M. 2020. Federated visual classification

with real-world data distribution//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 76-92 [DOI: 10.1007/978-3-030-58607-2_5]
Huang L, Yin Y F, Fu Z, Zhang S F, Deng H and Liu D B. 2020. LoAdaBoost: loss-based AdaBoost federated machine learning with reduced computational complexity on IID and non-IID intensive care data. PLoS One, 15 (4): #0230706 [DOI: 10.1371/journal.pone.0230706]
Kaissis G A, Makowski M R, Rückert D and Braren R F. 2020. Secure, privacy-preserving and federated machine learning in medical imaging. Nature Machine Intelligence, 2(6): 305-311 [DOI: 10.1038/s42256-020-0186-1]
Karimireddy S P, Kale S, Mohri M, Reddi S J, Stich S U and Suresh A T. 2020. SCAFFOLD: stochastic controlled averaging for federated learning//Proceedings of the 37th International Conference on Machine Learning. Virtual Event: JMLR.org: 5132-5143
Krizhevsky A. 2009. Learning Multiple Layers of Features from Tiny Images. University of Toronto
Kumar R, Khan A A, Kumar J, Zakria, Golilarz N A, Zhang S M, Ting Y, Zheng C Y and Wang W Y. 2021. Blockchain-federated-learning and deep learning models for COVID-19 detection using CT imaging. IEEE Sensors Journal, 21(14): 16301-16314 [DOI: 10.1109/JSEN.2021.3076767]
Lecun Y, Bottou L, Bengio Y and Haffner P. 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11): 2278-2324 [DOI: 10.1109/5.726791]
Li Q B, Diao Y Q, Chen Q and He B S. 2022. Federated learning on Non-IID data silos: an experimental study//Proceedings of the 38th IEEE International Conference on Data Engineering (ICDE). Kuala Lumpur, Malaysia: IEEE: 965-978 [DOI: 10.1109/ICDE53745.2022.00077]
Li X X, Jiang M R, Zhang X F, Kamp M and Dou Q. 2021a. FedBN: federated learning on Non-IID features via local batch normaliza-

- tion [EB/OL]. [2023-12-01]. <https://arxiv.org/pdf/2102.07623.pdf>
- Li Z X, Lin T, Shang X Y and Wu C. 2023. Revisiting weighted aggregation in federated learning with neural networks//Proceedings of the 40th International Conference on Machine Learning. Honolulu, USA: JMLR.org: 19767-19788
- Li Q B, He B S and Song D. 2021b. Model-contrastive federated learning//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 10708-10717 [DOI: 10.1109/CVPR46437.2021.01057]
- Li T, Sahu A K, Zaheer M, Sanjabi M, Talwalkar A and Smith V. 2020a. Federated optimization in heterogeneous networks [EB/OL]. [2023-12-01]. <https://arxiv.org/pdf/1812.06127.pdf>
- Li X, Huang K X, Yang W H, Wang S S and Zhang Z H. 2020b. On the convergence of FedAvg on Non-IID data [EB/OL]. [2023-12-01]. <https://arxiv.org/pdf/1907.02189.pdf>
- Liu Y, Huang A B, Luo Y, Huang H, Liu Y Z, Chen Y Y, Feng L C, Chen T J, Yu H and Yang Q. 2020. FedVision: an online visual object detection platform powered by federated learning//Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York, USA: AAAI: 13172-13179 [DOI: 10.1609/aaai.v34i08.7021]
- McMahan H B, Moore E, Ramage D, Hampson S and Arcas B A Y. 2017. Communication-efficient learning of deep networks from decentralized data//Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. Fort Lauderdale, USA: [s.n.]: 1273-1282
- Wang H Y, Yurochkin M, Sun Y K, Papailiopoulos D and Khazaeni Y. 2020a. Federated learning with matched averaging [EB/OL]. [2023-12-01]. <https://arxiv.org/pdf/2002.06440.pdf>
- Wang J Y, Liu Q H, Liang H, Joshi G and Poor H V. 2020b. Tackling the objective inconsistency problem in heterogeneous federated optimization//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: 7611-7623
- Wu H D and Wang P. 2021. Fast-convergent federated learning with adaptive weighting. IEEE Transactions on Cognitive Communications and Networking, 7 (4): 1078-1088 [DOI: 10.1109/TCCN.2021.3084406]
- Xia P P, Zhang L and Li F Z. 2015. Learning similarity with cosine similarity ensemble. Information Sciences, 307: 39-52 [DOI: 10.1016/j.ins.2015.02.024]
- Yuan Q Q, Shen H F, Li P X and Zhang L P. 2010. Adaptively regularized multi-frame image super-resolution reconstruction. Journal of Image and Graphics, 15 (12): 1720-1727 (袁强强, 沈焕锋, 李平湘, 张良培. 2010. 自适应正则化多幅影像超分辨率重建. 中国图象图形学报, 15 (12): 1720-1727) [DOI: 10.11834/jig.20101202]
- Yurochkin M, Agarwal M, Ghosh S, Greenewald K, Hoang T N and Khazaeni Y. 2019. Bayesian nonparametric federated learning of neural networks//Proceedings of the 36th International Conference on Machine Learning. Long Beach, USA: [s.n.]: 9252-9261
- Zhao Y, Li M, Lai L Z, Suda N, Civin D and Chandra V. 2022. Federated learning with Non-IID data [EB/OL]. [2023-12-01]. <https://arxiv.org/pdf/1806.00582.pdf>

作者简介

戚银城,男,教授,中国电机工程学会电力通信专委会副主任委员,主要研究方向为电力系统通信与信息处理。

E-mail: qiyech@ncepu.edu.cn

霍亚琳,女,硕士研究生,主要研究方向为联邦学习。

E-mail: hylin_2000@163.com

王宁,男,硕士研究生,主要研究方向为联邦学习。

E-mail: 2379861156@qq.com

侯禹,男,工程师,主要研究方向为电力物联网技术。

E-mail: 532144069@qq.com