计 算 机 研 究 与 发 展
Journal of Computer Research and Development
https://crad.ict.ac.cn
62(12): 3002−3012, 2025

非独立同分布数据下层次化 Sinkhorn 距离聚类联邦学习算法

赵 瑞 段小文 刘 新 周 睿 周庆国
（兰州大学信息科学与工程学院 兰州 730030）
（zhaor19@lzu.edu.cn）

Hierarchical Sinkhorn Distance-Based Clustering Federated Learning Algorithm for Non-IID Data

Zhao Rui, Duan Xiaowen, Liu Xin, Zhou Rui, and Zhou Qingguo
（*School of Information Science and Engineering*, *Lanzhou University*, *Lanzhou* 730030）

Abstract Federated learning, as an emerging distributed neural network training paradigm in edge computing, faces a key challenge of data heterogeneity over clients, which significantly impacts global model performance. While clustering federated learning, serving as a promising solution, has shown improvements in model accuracy, its effectiveness is limited by the insufficient utilization of local data distribution statistics. To address this issue, we propose a novel federated learning framework named HS-CFA (hierarchical Sinkhorn distance-based clustering federated algorithm), which is designed to optimize the aggregation weights of local models under the heterogeneous data environments, thereby enhancing the performance of the global model. It utilizes the entropy-regularized optimal transport cost to capture the characteristics of local data distribution and dynamically adjusts the aggregation weights of local models with a hierarchical clustering strategy. Specially, HS-CFA employs the Sinkhorn distance as a lightweight optimal transport cost metric to measure distributional dissimilarities across clients. Furthermore, it adopts a hierarchical two-layer clustering mechanism, combining density-based spatial clustering and average aggregation during the server training phase, facilitating the dynamic and adaptive adjustment of local model aggregation weights. Experimental results on multiple benchmark datasets demonstrate that HS-CFA significantly enhances the accuracy and robustness of the global model in scenarios with highly heterogeneous distribution settings.

Key words edge computing; federated learning; optimal transport; data heterogeneity; clustering optimization

摘 要 联邦学习作为一种边缘计算中的新兴分布式神经网络训练方法面临客户端数据异构性挑战，其中聚类联邦学习被认为是一种颇具潜力的解决方案，然而现有聚类联邦学习算法未深入探究量化客户端数据分布差异. 针对该问题提出了一种新颖的层次化 Sinkhorn 距离聚类联邦学习算法（hierarchical Sinkhorn distance-based clustering federated learning algorithm, HS-CFA），采用最优传输代价衡量局部训练时客户端数据分布特性，提出层次化聚类策略动态调整全局模型聚合时的参与权重. 具体而言，HS-CFA 引入 Sinkhorn 距离量化客户端间的分布差异，提出使用基于密度的聚类算法在服务器端进行动态层次聚类. 在多个基准数据集上的实验结果表明，相比传统算法，所提算法在高度数据分布异构性的场景中显著提升了全局模型的准确率和鲁棒性.

关键词 边缘计算；联邦学习；最优传输；数据异构性；聚类优化

收稿日期：2024-12-29；**修回日期**：2025-04-01
基金项目：国家重点研发计划项目（2023YFB4503903）
This work was supported by the National Key Research and Development Program of China (2023YFB4503903).
通信作者：周庆国（zhouqg@lzu.edu.cn）

中图法分类号 TP393

DOI: 10.7544/issn1000-1239.202440969 CSTR: 32373.14.issn1000-1239.202440969

随着物联网设备的广泛使用，大量碎片化数据的产生推动了分布式机器学习的研究发展[1]. 传统的数据中心采用集中式架构，即数据中心汇总处理经由物联网设备采集传输的数据，并依靠自身计算资源训练神经网络模型. 然而在有限通信资源下，该架构会因海量分布式数据的传输导致高延迟问题. 其次数据中心作为单点存储和计算节点，在遭遇网络攻击时，可能存在大规模的隐私数据泄露风险[2]. 更严重的是一旦服务器模型推理异常，导致物联网智能应用的全面瘫痪[3]，影响所有连接的边缘设备应用.

针对上述问题，边缘计算[4-5]提出将数据处理存储前置到物联网边缘节点而非集中式在数据中心处理的优化方案，从而提高数据处理效率. 然而由于神经网络模型学习依赖大量训练数据，而边缘设备通常面临数据不足和计算能力受限[6]问题，导致单一边缘设备节点难以训练高性能模型. 为了进一步应对该挑战，2016 年谷歌团队提出了分布式机器学习方法——联邦学习（federated learning, FL）[7]，该训练范式允许多个边缘节点[8-9]作为客户端利用本地数据协作训练目标模型而无需共享原始数据，通过上传目标模型梯度或参数在服务器端采用聚合操作进行模型更新. 该方法不仅有效减少了传统分布式训练因数据传输带来的通信开销，同时也在保障边缘设备数据隐私的基础上，提升了目标模型性能.

然而实际应用场景中，联邦学习往往难以满足其基础假设，即各客户端的本地训练数据应保持独立同分布特性. 例如，边缘节点在数据采集过程中，由于设备的使用环境、用户行为模式以及数据来源的不同，其数据在分布、标签以及规模上可能存在显著差异，这导致本地数据呈现非独立同分布（non-independent and identically distributed, non-IID）特性[10]. 这一特性会引发联邦学习中目标模型训练过程收敛缓慢、模型泛化性能不佳等问题. 因此数据异构性是联邦学习面临的主要挑战，一方面客户端由于自身数据有限，难以单独训练高性能模型，需依赖多个客户端协作学习；另一方面，与数据分布差异较大的客户端进行协作时，目标模型泛化精度难以得到保证. 针对该问题，已有相关研究提出了多种应对策略[11-15]，其中聚类联邦学习（clustered federated learning, CFL）[16-21]被认为是一种颇具潜力的算法，通过对相似数据分布的客户端模型聚类，减少分布差异性，提升聚合时目标模型的泛化能力. 尽管现有 CFL 研究取得了诸多进展，但相关研究未充分考虑训练时聚合模型的动态性，从而假设聚类簇的数量为固定的先验知识，缺乏自适应调整机制. 其次对于数据分布差异这一关键度量，现有算法对其缺乏深入研究，难以精准量化客户端数据分布异构性.

本文提出了一种面向 non-IID 分布的层次化 Sinkhorn 距离聚类联邦学习算法（hierarchical Sinkhorn distance-based clustering federated learning algorithm, HS-CFA）. 该算法挖掘各客户端的数据分布差异，通过 2 层聚类机制提升联邦学习全局模型的性能. 具体而言，HS-CFA 将数据分布差异视为分布之间的最优传输代价，采用熵正则化的轻量 Sinkhorn 距离[22]将该代价量化为客户端的相似度度量. 基于该度量，HS-CFA 提出了一种高效简单的聚类策略，利用 DBSCAN（density-based spatial clustering of applications with noise）聚类算法[23]将具有相似数据特征的本地模型聚类为簇，簇内采用平均聚合生成簇模型. 随后对簇模型聚类生成全局模型. 通过该聚类策略，HS-CFA 实现了自适应调节本地模型聚合权重，有效提升了全局模型的泛化能力，加速联邦学习的收敛过程.

本文主要贡献有 3 点：

1）基于概率分布的最优传输[24]理论，HS-CFA 引入 Sinkhorn 距离作为熵正则化最优传输代价，衡量本地数据分布之间的差异，从而解决了服务器端无法有效表征客户端数据分布特性的问题，同时降低了最优传输代价的计算复杂度.

2）设计了服务器端双层聚合机制，在无需先验聚类知识的条件下，利用客户端本地训练时与全局模型的相似度度量进行聚类，同时排除离群本地模型带来的干扰，该机制能动态平衡本地模型聚合权重.

3）在高度 non-IID 数据异构场景中对 HS-CFA 进行了大量实验，结果表明 HS-CFA 在目标模型精度与鲁棒性方面均优于现有的聚类联邦学习算法，在高数据异构场景中表现尤为突出.

1 相关工作

联邦学习是一种新兴的分布式机器学习方法，由 Mcmahan 等人[25]提出的 FedAvg 算法是该领域的奠基性工作，该算法在服务器端训练阶段对上传的

本地模型进行加权平均聚合,以迭代更新全局模型.然而在面对 non-IID 数据分布时,FedAvg 算法的精度会显著下降,甚至出现模型无法收敛的问题.文献 [26] 进一步指出该问题的根源在于本地模型优化目标与全局模型优化目标的不一致性.

为解决这一问题,近年相关研究提出了多种轻量级优化算法,根据优化目标可分为 2 类:

一类是为了缓解本地模型数据分布导致的不一致性,对客户端训练过程的目标损失函数优化.例如,FedProx[27] 和 SCAFFOLD [28] 通过在本地模型的目标函数中加入正则惩罚项来对齐全局模型和本地模型;FedCM[29] 和 FedSpeed[30] 则利用乘子交替方向法[31] 在本地模型的损失函数中加入约束条件.

另一类算法则在服务器端聚合本地模型时加入聚类算法从而优化聚合权重分配策略.Sattler 等人[20] 首次提出 CFL 算法,通过迭代二分法将具有相似数据分布的本地模型进行分簇,提高簇内模型的性能,这种算法显著减轻了数据异质性的负面影响,从而提高了模型的准确性和收敛速度.Ghosh 等人[16] 提出 IFCA 算法预先设定聚类簇,增强了联邦学习簇模型个性化能力.FLSC[19] 采用松弛聚类的算法,在 IFCA 算法的基础上对聚类不进行严格划分,提升了个性化联邦学习效果.然而 IFCA 和 FLSC 性能依赖于多个个性化簇模型,因此在生成具备泛化能力的单一全局模型方面存在局限性.分层联邦学习算法 FL-HC[17] 组织客户端进行分层聚类来提高模型的准确性;文献 [21] 提出了 FLIS 算法,利用服务器端的代理数据集测量本地模型的相似度从而优化聚合权重.分析现有聚类联邦学习算法,优化的关键点集中在聚类簇的划分策略以及聚类相似度度量指标的选择上.

因此与现有研究的区别在于,本文研究聚焦单一目标模型的聚类联邦学习算法,为 non-IID 场景下提供了一种双层聚类优化策略.同时与现有研究方法中采用模型损失、欧几里得距离或余弦距离[32] 来衡量客户端数据分布的差异不同,本文利用熵正则化的最优传输代价动态衡量数据分布相似度,进一步为了降低计算复杂度,提出使用轻量化的 Sinkhorn 距离进行高效计算,从而在保证相似度度量精度的同时,提升全局模型的训练效率和收敛性能.

2 预备知识

2.1 联邦学习

经典联邦学习算法[25] 用于分布式协作训练一个全局泛化模型.算法可以概括为在一个全局训练周期中进行 4 个步骤:1)服务器将全局模型分发至客户端;2)客户端利用本地数据训练;3)客户端将训练后的模型参数或梯度上传至服务器;4)服务器通过加权聚合的方式进行全局训练.通过不断迭代全局训练周期直至模型收敛.

算法 1. 联邦学习算法流程.

输入:全局初始化模型θ_*^0,客户端总数N,客户端与服务器通信轮次T,客户端训练轮次K;

输出:全局模型θ_*^T.

① 初始化全局模型参数θ_*^0;

② for $t\leftarrow 0$ to T do

③ 随机选择包含M个客户端集合;

④ 发送 θ_*^0到M;

⑤ 每个客户端$i\in M^t$并行执行;

⑥ 使用本地数据集进行E次训练;

⑦ 更新客户端模型参数θ_i^t至服务器端;

⑧ 服务器执行$\theta_*^{t+1}=\dfrac{1}{|M|}\sum\limits_{i\in M}\theta_K^t$;

⑨ end for

⑩ 返回θ_*^T.

算法 1 描述联邦学习的基础训练范式,可概括为客户端训练周期和服务器训练周期.客户端训练周期中,客户端模型利用本地数据训练,其目标优化函数表示为

$$F_i(\theta)=\mathbb{E}_{(x,y)\sim D_i}\left[\ell(\theta;x,y)\right], \tag{1}$$

其中$\ell(\theta;x,y)$是损失函数,D_i是第i个客户端的数据分布,$\mathbb{E}$表示对客户端数据分布的期望.

在服务器训练周期中,由于无法获得客户端数据,全局模型只能采用间接训练的方式,通常采用最小化全局目标函数$F(\theta)$,该目标函数是所有客户端数据分布的总体损失函数,其表示为:

$$F(\theta)=\frac{1}{|S_i|}\sum_{z=1}^{S_i}\mathbb{E}_{(x,y)\sim D_i}\left[\ell(\theta;x_z,y_z)\right], \tag{2}$$

其中S_i是第i个客户端的样本数据数量,(x_z,y_z)则是第z个样本.事实上联邦学习服务器采用的训练方式是分布式随机梯度下降算法,将不同客户端上传的模型看待为大小不一的批处理训练结果,在此基础上进行加权期望,权重为不同客户端训练数据集大小.然而分布式随机梯度下降算法的前提是数据由服务器进行采样分发,因此分布式数据通常满足独立同分布特性,因而分布式训练局部优化目标一致.但在联邦训练范式中,分布式数据是客户端自有数据,不

同客户端数据之间存在异构性：1)数据量分布偏差，即相同标签的数据数量存在显著差异；2)标签分布偏差，数据标签在客户端之间分布不一致. 这种 non-IID 的特性导致客户端分布式训练时的优化目标存在差异，进一步在服务器利用分布式随机梯度下降进行聚合时，会显著影响全局模型泛化性能，导致客户端数据异构性越强，全局模型泛化精度越低.

2.2 最优传输理论

最优传输理论[22,24]提供了一种在 2 个概率测度之间进行操作的数学方法，即通过定义代价函数，如式(3)所示.

$$c: X\times Y\to \mathbb{R}\cup\{\infty\}. \tag{3}$$

最优传输希望找到一个传输代价使得从$x\in X$转移到$y\in Y$的总传输代价最小，$c(x,y)$表示将$x\in X$转移到$y\in Y$的代价，如式(4)所示.

$$\min_T M(T)=\int_X c(x,T(x))\mathrm{d}\mu(x), \tag{4}$$

其中概率测度$\mu\in P(X)$，概率测度$v\in P(Y)$，因此$v=T\times\mu$是从$x\in X$转移到$y\in Y$的测度. 最优传输理论中的 Wasserstein 距离，即是量化一个概率分布转换为另一个分布所需的最优传输代价. 该测度被广泛使用到了深度学习的生成网络模型损失函数设计中，度量生成样本与已有样本之间的相似度，对样本空间的数据分布进行学习.

然而直接在较大规模的数据分布上计算 Wasserstein 距离代价较高，通常引入熵正则化项减少计算成本：

$$objective(\pi)=\int_{X\times Y} c(x,y)\mathrm{d}(x,y)+\gamma Ent(\pi), \tag{5}$$

其中π为最优传输代价，γ为控制权重，熵正则化项$Ent(\pi)$定义为

$$Ent(\pi)=\int_{X\times Y}\log\frac{\mathrm{d}\pi(x,y)}{\mathrm{d}\mu(x)\mathrm{d}v(y)}\mathrm{d}\pi(x,y). \tag{6}$$

为了进一步加快计算，对目标函数对偶化得到 Sinkhorn 距离[22]，这是一种快速计算求解带有熵正则化的最优传输问题的迭代方法，利用矩阵的归一化迭代逼近最优传输测度.

$$\begin{aligned}Sinkhorn(\mu,v)=&\inf_{(\mu,v)}\int_{X\times Y}c(x,y)\mathrm{e}^{\mu(x)+v(y)}\mathrm{d}\mu(x)\mathrm{d}v(y)-\\&\frac{1}{\gamma}\left(\int_X\mathrm{e}^{\mu(x)}\mathrm{d}\mu(x)-1\right)-\\&\frac{1}{\gamma}\left(\int_Y\mathrm{e}^{v(y)}\mathrm{d}v(y)-1\right).\end{aligned} \tag{7}$$

Sinkhorn-Knopp[20]算法通过对μ和v交替更新的方式，最终得到一个收敛的$Sinkhorn(\mu,v)$，更新计算公式如下：

$$\mu^{k+1}=\log\left(\frac{1}{\mu(x)}\int_X\mathrm{e}^{v^k(y)-c(x,y)}\right)\mathrm{d}v(y), \tag{8}$$

$$v^{k+1}=\log\left(\frac{1}{v(x)}\int_X\mathrm{e}^{\mu^{k+1}(x)-c(x,y)}\right)\mathrm{d}\mu(x). \tag{9}$$

3 层次化 Sinkhorn 距离聚类联邦学习算法

联邦学习客户端数据私有且 non-IID 特性使得其存在 2 个显著限制：1)服务器端无法获取客户端数据进行计算；2)客户端无法获取其他客户端模型信息进行优化. 而本文提出基于 Sinkhorn 距离的层次化聚类联邦学习算法有效缓解上述问题并优化模型聚合过程.

如图 1 中间区域所示，HS-CFA 动态评估客户端数据分布的统计特性，采用最优传输代价来量化 non-IID 客户端数据分布. 在每个全局通信轮次中，客户

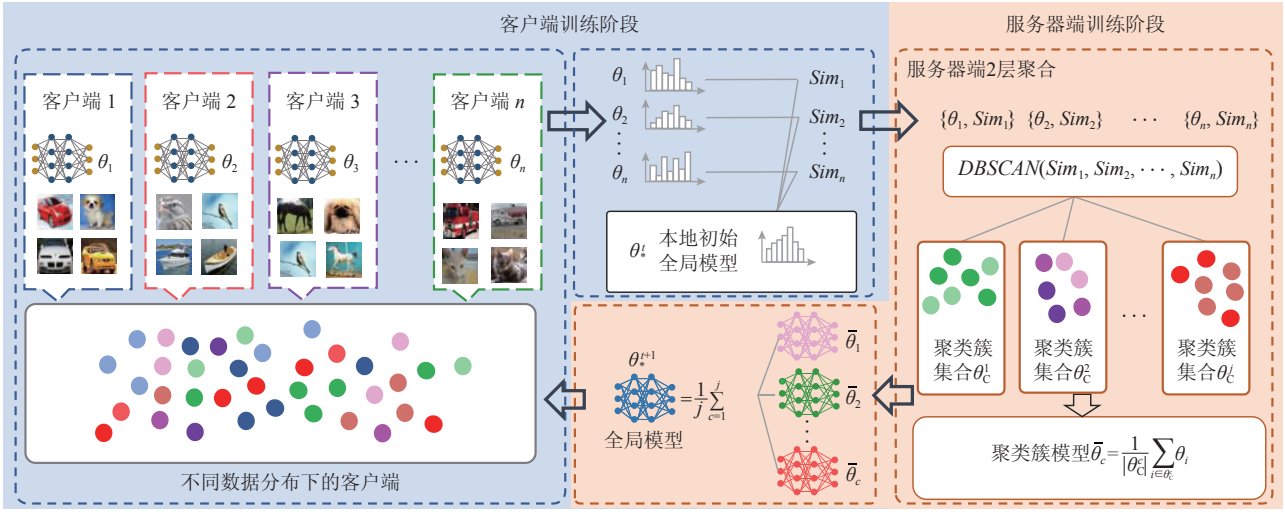


Fig. 1 Schematic diagram of HS-CFA algorithm

图 1 HS-CFA 算法示意图

端利用 Sinkhorn 距离计算客户端模型与全局模型之间的差异, 客户端上传本地模型及相似性度量至服务器. 如图 1 右侧区域所示, 服务器端间接利用本地模型与全局模型的相似度度量描述各客户端数据分布之间的差异性. 使用层次化聚类方式进行聚合, 第 1 层聚类利用 DBSCAN 算法基于相似度度量生成聚类簇模型; 第 2 层采用平均聚合对聚类簇模型聚类, 生成全局模型. 随后服务器将聚类簇模型和全局模型下发至客户端进行迭代训练.

3.1 客户端训练

HS-CFA 在客户端训练周期中通过度量本地模型与全局模型之间的差异性, 间接地在不违反数据隐私的前提下向服务器提供客户端数据分布特性.

算法 2. HS-CFA 客户端训练.

输入: 全局模型 θ_*, 聚类簇模型 $\bar{\theta}_c^t$, 聚类簇数 j, 局部数据 $(x,y)\sim D_i$;

输出: 本地模型 θ_i, 数据分布相似度 Sim_i, 客户端标识 i.

/*T为客户端训练和全局模型通信轮次;

K为客户端训练轮次;

η为客户端模型训练学习率;

F_i为客户端训练损失函数. */

① for $t\leftarrow 0$ to T do

② if θ_c 不存在

③ $\bar{\theta}_c^t=\theta_*^t$, $\theta_i^t=\theta_*^t$;

④ else

⑤ $\theta_i^t=\theta_*^t-\dfrac{1}{j-1}\left(\bar{\theta}_c^t\right)$;

⑥ end if

⑦ for $k\leftarrow 1$ to K do

⑧ $\theta_{i,k+1}^t\leftarrow\theta_{i,k}^t-\eta\nabla F_i(\theta|(x,y)\sim D_i)$;

⑨ end for

⑩ $Sim_i=Sinkhorn\left(\theta_{i,K}^t,\theta_*^t|(x,y)\sim D_i\right)$;

⑪ $\{\theta_{i,K}^t,Sim_i,i\}$发送至服务器;

⑫ end for

算法 2 描述了图 1 中 HS-CFA 客户端训练过程. 考虑到本地模型对于全局数据分布未知, 初始模型 θ_*^t 对所有客户端数据分布具备相同的度量尺度, 因此假设初始全局模型和客户端经过 K 轮训练后的模型是 2 个函数表示, 则对局部数据分布采样计算得到相关函数的值域分布, 可间接地表示客户端数据分布对于全局模型和本地模型的差异性. 具体在联邦学习的每一轮中, 客户端通过对训练局部轮次 K 后产生的本地模型 $\theta_{i,K}^t$ 和初始全局模型 θ_*^t 根据式(8)(9)交替迭代计算 Sinkhorn 距离, 获得相似度度量 Sim_i.

此外与其他算法相比, HS-CFA 要求服务器传输: 全局模型 θ_* 和该客户端所在的聚类簇模型 $\bar{\theta}_c$ 两个模型以及全局聚类时产生的聚类簇数量 j.

如果客户端首次被选中参与训练, 则本地模型初始化与全局模型保持一致, 即使用聚类簇模型进行初始化. 而对于已参与过联邦训练的客户端, 其本地模型的初始化方式不同于传统联邦学习算法直接使用 θ_*, 而是按照算法 2 中的行⑤进行更新.

HS-CFA 利用该方式对本地模型训练初始化, 间接获取其他聚类簇的全局模型数据分布先验知识, 从而让算法避免陷入局部最优解, 排除自身聚类簇模型 $\bar{\theta}_c$ 的影响.

3.2 服务器端训练

根据预备知识中的联邦学习数据异构性来源于全局模型的优化目标无法对齐的性质, 本文从 VC 理论分析联邦学习全局模型的泛化上限.

定理 1[13]. 假设 H 为一个 VC 维为 d 的假设空间, $\hat{D}_{S_i}$ 代表联邦学习从每个源域中抽取大小为 m 的样本所产生的先验分布, $\hat{D}_T$ 为目标域数据分布, 因此存在 $\alpha_i\in\mathbb{R}^+$ 且 $\sum_{i=1}^{N}\alpha_i=1$, 那么对于假设空间中的函数 $h\in H$, 至少有 $1-\xi$ 的概率存在:

$$\epsilon_T(h)\leqslant\hat{\epsilon}_{S_i}\left(\sum_i\alpha_i h_{S_i}\right)+\sum_i\alpha_i\left(\frac{1}{2}\hat{d}_{H\times H}\left(\hat{D}_{S_i},\hat{D}_T\right)\right)+\sigma_{vc}+\lambda, \tag{10}$$

其中 σ_{vc} 是 VC 理论泛化误差常量, λ 代表最低误差之和的无穷小量, $\hat{d}_{H\times H}(\hat{D}_{S_i},\hat{D}_T)$ 代表客户端和全局数据分布之间的散度. 定理 1 表明全局模型的泛化上限一方面由本地模型泛化能力 $\epsilon(h_{S_i})$ 决定, 另一方面则受全局模型的数据分布与局部模型的数据部分差异性的影响, 考虑差异性客观存在, 无法消除 $\hat{d}_{H\times H}\left(\hat{D}_{S_i},\hat{D}_T\right)$ 的影响. 但是客户端权重 α_i 不仅影响本地模型聚合性能, 同时可动态地调整客户端与数据分布之间的差异. 因此为了尽可能最大化全局模型泛化上限, 需要动态调整权重 α_i, 提高对全局模型分布有益的客户端的权重, 同时限制对全局模型数据分布有害的客户端的权重.

考虑到经典联邦学习算法中服务器训练实质是对客户端模型进行加权聚合以模拟分布式随机梯度下降法. 权重 α_i 被设置为 $|D_i|/\left|\sum_{i=1}^{n}D_i\right|$, 即客户端 i 数据量与所有客户端数据量之比. 本文提出在服务器训练周期中采用 DBSCAN 算法和层次化聚类方法训

练全局模型的策略，希望客户端的权重α_i的设置不再依赖数据集的大小，而基于对全局数据分布的公平化差异性贡献. 假设全局数据分布 $D\in\{D_1,D_2,\cdots,D_n|n\in M\}$，其中 M是参与联邦学习的客户端的总数. 客户端上传的Sim_i相似度可度量不同的$\{D_1,D_2,\cdots,D_n\}$数据分布特性，通过相似度对客户端模型划分聚类，会使相似本地数据分布的客户端模型聚类为簇. 考虑到不同聚类簇包含的客户端模型数目不一，需要平衡不同聚类簇对于全局模型的数据分布的贡献，因此 HS-CFA 均采取平均聚合策略.

算法 3. HS-CFA 双层聚合算法.

输入：参与训练的客户端模型$\{\theta_1,\theta_2,\cdots,\theta_i\}$，参与训练的客户端模型与全局模型的相似度$\{Sim_1,Sim_2,\cdots,Sim_i\}$，客户端标识$i$；

输出：全局模型θ_*，聚类簇模型集合θ_{Cluster}，聚类簇数目j.

/* T为客户端训练时和全局模型通信轮次；

$\mathcal{N}_\epsilon$为 DBSCAN 算法的领域范围；

$\min_s$为 DBSCAN 算法的最小采样量. */

① for $t\leftarrow 0$ to T do

② if $t=0$

③ 初始化θ_*^0, $\theta_{\text{Cluster}}=\varnothing$;

④ end if

⑤ for M客户端 do

⑥ 接收$\theta_M=\{\{\theta_i,Sim_i,i\},i\in m\}$;

⑦ end for

⑧ $\left\{\theta_{\text{C}}^1,\theta_{\text{C}}^2,\cdots,\theta_{\text{C}}^j\right\}=DBSCAN(\theta_M,\mathcal{N}_\epsilon,\min_s)$;

⑨ for $\theta_{\text{C}}^c\in\left\{\theta_{\text{C}}^1,\theta_{\text{C}}^2,\cdots,\theta_{\text{C}}^j\right\}$ do

⑩ $\bar{\theta}_c=\dfrac{1}{|\theta_{\text{C}}^c|}\sum\limits_{i\in\theta_{\text{C}}^c}\theta_i$; /*第 1 次簇内聚合*/

⑪ $\theta_{\text{Cluster}}\leftarrow\theta_{\text{Cluster}}\cup\bar{\theta}_c$;

⑫ end for

⑬ $j=|\theta_{\text{Cluster}}|$;

⑭ $\theta_*^{t+1}=\dfrac{1}{j}\sum\limits_{c=1}^{j}\bar{\theta}_c$; /*第 2 次簇间聚合*/

⑮ $\theta_{\text{Cluster}}=\varnothing$;

⑯ 下发$\{\theta_*^{t+1},\bar{\theta}_c,j\}$至客户端；

⑰ end for

算法 3 包含了服务器训练周期的 2 个阶段.

1）初始化阶段. 服务器初始用于聚合时保存t轮次的聚类簇模型集合.

2）聚类训练阶段. 服务器端在接收到所有客户端发送的$\{\theta_i,Sim_i,i\}$数据后，基于 DBSCAN 算法按照每个客户端的相似性度量Sim_i进行迭代聚类，该过程将数据分布相似的客户端划分至相同的簇内，以确保簇内客户端模型在数据分布上具有较高的一致性. 对于每个聚类簇，服务器采用平均聚合的方式生成聚类簇模型，并进一步对这些聚类簇模型进行 2 次聚类，最终得到全局模型. 考虑到全局数据分布的复杂性，HS-CFA 通过 DBSCAN 不需要设定聚类簇的先验知识的特点，一方面尽可能将相似度量统计值的局部客户端模型进行聚合，另一方面自动过滤掉与聚类簇分布差异过大的离群本地模型. 从而确保 2 次聚类后得到的全局模型排除离群客户端数据分布的影响，有效衡量各数据分布下本地模型对全局模型聚合的贡献.

在获得全局模型后，服务器需更新该模型至客户端. 鉴于不同轮次中服务器选取的客户端可能存在差异，分发阶段采用以下策略：对于上一轮已参与训练且在当前轮次中再次被选中的客户端，服务器根据客户端标识i发送聚类簇模型$\bar{\theta}_c$、全局模型θ_*和聚类簇数目j至该客户端. 对于首次被选中的客户端，服务器仅传输全局模型θ_*，确保模型初始化的一致.

4 实验结果与分析

4.1 实验数据设置

为了验证 HS-CFA 算法的有效性，本文选择了 3 个评估联邦学习的基准数据集：Fashion-MNIST[33]，CIFAR-10[34]，CIFAR-100[34]. Fashion-MNIST 包含 10 个类别的灰度图像，每个类别有 6 000 张训练图像和 1 000 张测试图像，图像分辨率为 28×28 像素；CIFAR-10 包含 10 个类别的彩色图像，每个类别有 5 000 张训练图像和 1 000 张测试图像，图像分辨率为 32×32 像素；CIFAR-100 包含 100 个类别的彩色图像，每个类别有 500 张训练图像和 100 张测试图像，图像分辨率为 32×32 像素.

数据集均采用 2 种数据异构性设置：1）通过狄利克雷分布$Dir(\delta)$[35]对数据集进行划分，其中δ用于衡量数据的异构性，系数值越小，数据分布的不均匀程度越高，即数据越偏离独立同分布. 2）基于文献 [14, 17]，对数据集中的图像进行随机 4 种角度旋转 (0°，90°，180°，270°). 旋转增强图像根据比例β混入客户端数据集.

4.2 实验设置

本文采用 2 种神经网络结构作为骨干网络：1）用于 Fashion-MNIST 数据集的 2 层多层感知机网络结构；2）用于 CIFAR-10 和 CIFAR-100 数据集的 AlexNet[36]

网络.

训练参数设置采用随机梯度下降算法, 优化器动量设置为 0.9, 权重衰减为 0.000 1, 学习率为 0.01. 批处理大小设置为 32(对于 Fashion-MNIST)和 64(对于 CIFAR-10 和 CIFAR-100).

联邦学习参数设置中客户端数目M为 50, 每个全局交互轮次中参与度P为 80%. 每个客户端的局部训练次数T为 5, 全局交互轮次C为 100 轮. 数据异构性方面, 狄利克雷分布设置δ为 0.1 和 0.3. 旋转增强混入β为 30% 和 50% 的数据. 在 DBSCAN 算法中邻域阈值N_ϵ设置为 0.3, 最小样本数$\min_s$设置为 5. 客户端计算 Sinkhorn 采样设置为 16. 考虑到相似度用于 DBSCAN 聚类计算无需过于精准, 为尽可能降低客户端的计算负担, Sinkhorn 的迭代次数设置为 5.

所有实验均在 Ubuntu 20.04 操作系统下, 通过 32 GB 的 Nvidia Tesla V100 服务器使用 PyTorch[37] 框架进行训练与测试, 并根据不同的随机种子重复运行了 5 次.

4.3 实验结果分析

本文选取了基线算法 FedAvg[25]; 3 种客户端优化联邦学习算法, 包括 FedProx[27], FedRobust[38], SCAFFOLD[28]; 2 种针对单一目标模型的聚类联邦学习 FL-HC[17] 和 FLIS[21]; 2 种个性化联邦聚类算法 IFCA[16] 和 FLSC[19]. 由于其会产生多个全局模型, 测试选择了性能最优的个性化模型作为基线.

表 1 展示了在数据异构性场景下, Fashion-MNIST 数据集的全局模型平均测试准确率, 其中异构性由狄利克雷分布参数δ控制, 取值分别为 0.1 和 0.3. 在参数为 0.1 的情况下, 所有算法的测试准确率均低于参数为 0.3 下的情况, 同时测试准确率方差更大, 表明在极端 non-IID 场景中, 不同算法性能鲁棒性均有所下降. 相比其他基线算法, HS-CFA 的方差更小. 同时 HS-CFA 算法在参数为 0.1 取得了最优测试准确率, 在参数为 0.3 时, 其测试准确率略低于 FLIS, 但稳定性优于 FLIS. 这主要归因于 HS-CFA 采用的 DBSCAN 聚类算法能够筛除离群点, 从而降低在聚合全局模型时所产生的负面影响. 总体而言, HS-CFA 在不同 non-IID 设置下测试准确率变化较小, 表明 HS-CFA 在处理数据分布异构性方面具有良好的鲁棒性.

Table 1 Test Accuracy on Fashion-MNIST Under Different Dirichlet Coefficients

表 1 不同狄利克雷系数下在Fashion-MNIST的测试准确率 %

算法	δ=0.1	δ=0.3
FedAvg[25]	69.91±1.83	82.25±0.91
FedProx[27]	70.13±1.75	82.07±0.88
FedRobust[38]	71.04±1.34	83.14±0.67
SCAFFOLD[28]	72.23±1.22	84.17±0.59
IFCA[16]	71.52±1.06	84.02±0.49
FLSC[19]	72.07±0.83	84.88±0.41
FL-HC[17]	71.08±0.90	83.65±0.52
FLIS[21]	72.21±0.92	85.13±0.47
HS-CFA (本文)	72.79±0.81	85.07±0.36

注: "±"前数值表示平均值, "±"后数值表示标准差.

表 2 展示了旋转混入数据异构性设置下的测试结果, 其中β表示混入旋转数据的比例. 结果显示, 在客户端中混入旋转图片对所有算法的测试准确率影响较小, 各算法的准确率方差均较小. 在β=30% 时, 虽然 HS-CFA 的准确率略低于 SCAFFOLD, FL-HC, FLIS 算法, 但准确率差距均在 1% 以内, 同时其稳定性优于这 3 种算法. 在β=50% 时, HS-CFA 的准确率仅次于 SCAFFOLD, 但其稳定性仍优于 SCAFFOLD. 这表明, 在数据异构性水平较低的情况下, HS-CFA 算法依然展现出较强的竞争力.

Table 2 Test Accuracy on Fashion-MNIST Under Different Rotation Mixing Coefficients

表 2 不同旋转混合系数下在Fashion-MNIST的测试准确率 %

算法	β=30%	β=50%
FedAvg[25]	90.49±0.32	91.79±0.29
FedProx[27]	89.91±0.15	91.81±0.12
FedRobust[38]	90.25±0.08	92.33±0.07
SCAFFOLD[28]	91.84±0.87	93.56±0.64
IFCA[16]	91.21±0.92	92.67±0.70
FLSC[19]	90.64±0.17	92.30±0.11
FL-HC[17]	91.77±0.53	92.59±0.25
FLIS[21]	91.82.±0.65	92.49±0.32
HS-CFA (本文)	91.76±0.09	92.80±0.13

注: "±"前数值表示平均值, "±"后数值表示标准差.

图 2 展示了在 CIFAR-10 数据集中不同δ设置下相关算法测试准确率曲线. 实验表明, 在数据极端不平衡的条件下, HS-CFA 算法表现出最佳的性能. 相较于同样优化单一目标模型的聚类联邦学习算法 FL-HC 和 FLIS, HS-CFA 每轮训练更加平滑. 此外, HS-CFA 在大约 30 轮时便已实现较好的收敛性, 显示了聚类中利用 Sinkhorn 距离衡量客户端数据分布的相似性具有良好的效果, 并且能够有效地反映不同客户端数据分布的差异性.

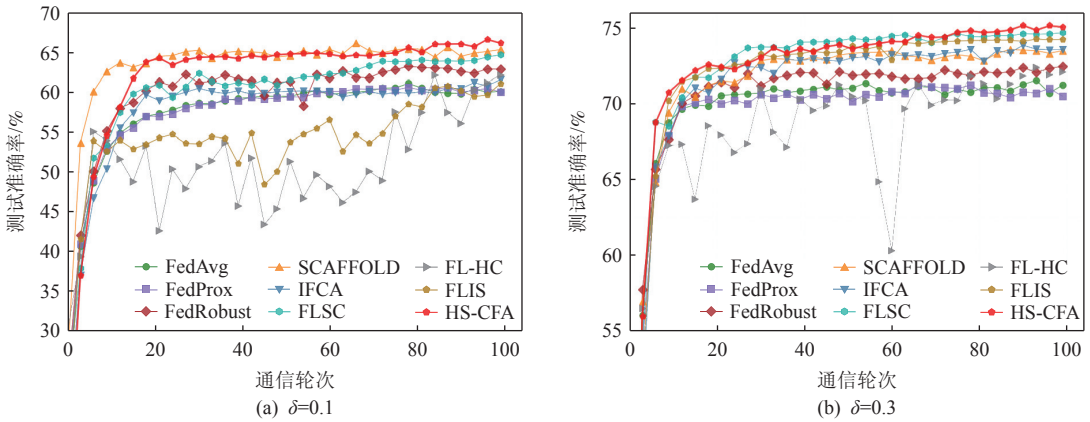


Fig. 2 Test accuracy curves of HS-CFA and baseline algorithms with different Dirichlet coefficients on CIFAR-10 dataset

图 2 CIFAR-10 数据集上不同狄利克雷系数下 HS-CFA 与基准算法测试准确率曲线

表 3 展示了在 CIFAR-100 数据集上，采用 δ=0.3 和旋转混入 β=30% 情况下，各算法的测试准确率. 由于 CIFAR-100 数据集中有 100 个类别，因此每个类别在客户端中的数据量相对较少，这进一步加剧了数据的不平衡性. 结果表明，HS-CFA 算法在测试准确率上仍然优于基准算法，并表现出较好的稳定性. 尽管其稳定性略低于 FLSC 和 IFCA，这一差异主要源于 IFCA 和 FLSC 生成个性化全局模型每轮记录的均是最优的个性化模型准确率，而 HS-CFA 侧重于训练单一全局模型.

Table 3 Test Accuracy on CIFAR-100 Under Different Dirichlet Coefficients and Rotation Mixing Coefficients

表 3 不同狄利克雷系数和旋转混合系数下在 CIFAR-100 的测试准确率 %

算法	δ=0.3	β=30%
FedAvg[25]	53.42±2.01	71.89±1.31
FedProx[27]	53.73±1.78	72.07±1.07
FedRobust[38]	55.78±1.27	73.14±0.59
SCAFFOLD[28]	56.41±1.58	74.33±0.62
IFCA[16]	49.93±0.76	71.02±0.79
FLSC[19]	53.15±0.81	75.18±0.40
FL-HC[17]	52.60±1.01	74.29±0.52
FLIS[21]	54.29±0.97	75.33±0.35
HS-CFA（本文）	55.72±0.84	75.31±0.16

注: “±”前数值表示平均值, “±”后数值表示标准差.

从 β=30% 的实验设置来看，旋转混入方式引入的数据异构性对测试准确率的影响较小，各算法的测试准确率差异不显著. 总体而言，基于聚类的联邦学习算法在测试准确率上普遍优于通过调整局部客户端目标函数的算法. 这说明，在分类任务中，随着标签类别数量的增加，聚类联邦学习通过优化客户端模型的权重分配策略，更有效地缓解了数据异构性带来的影响，从而提升了模型的整体性能.

表 4 展示了在 CIFAR-10 数据集上采用基本实验设置时 HS-CFA 各模块的计算时间开销. 其中本地训练设置为 5 轮，Sinkhorn 距离计算的采样样本为 16，第 1 层聚合则采用 DBSCAN 对 40 个客户端进行聚类.

Table 4 Processing Time of Each Module of HS-CFA on CIFAR-10 Dataset

表 4 HS-CFA各个模块在CIFAR-10数据集的处理时间 s

本地训练	相似度计算	第 1 层聚合	第 2 层聚合
0.512 45	1.133 26	0.474 51	0.080 35

尽管 HS-CFA 采用了轻量化的 Sinkhorn 计算方法来间接衡量本地数据的相似性，然而相似度计算时间仍然占据约 50% 的总时间.

相比之下，DBSCAN 算法在聚类时仅使用相似度度量，因此处理时间较少. 如果在该模块中引入空间索引结构等优化措施，时间性能将进一步提升.

4.4 消融实验

本文对 HS-CFA 中采用的 DBSCAN 方法的超参数进行了消融实验. 实验使用的数据集为 CIFAR-10，图 3 展示了 0.1, 0.3, 0.45, 0.6 这 4 种密度参数下全局模型的测试准确率.

通常情况下邻域半径为 0.3 时模型效果最佳. 然而训练达到 80 轮后，邻域半径为 0.6 下模型表现更优. 这一现象是由于随着训练轮次的增加，客户端数据分布的差异性持续存在，导致本地模型与全局模型的对齐难度加大. 因此，在后期训练阶段，聚类时应需要选择更大的邻域范围.

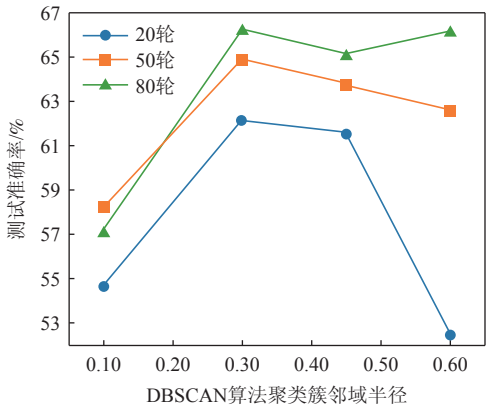


Fig. 3 Effect of different neighborhood radius of DBSCAN on test accuracy of global model

图 3 DBSCAN 中不同邻域半径对于全局模型测试准确率的影响

图 4 展示了在固定邻域半径的情况下, 选择簇内最小样本数时, HS-CFA 算法的测试准确率.

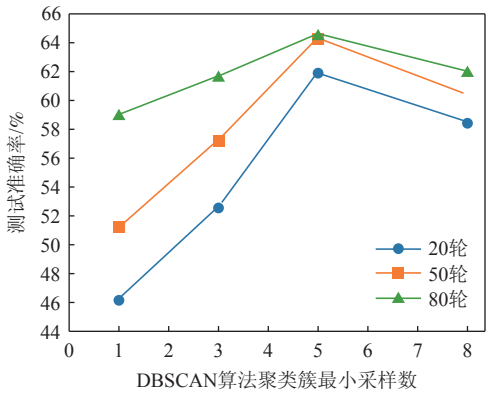


Fig. 4 Effect of different minimum sample amounts of DBSCAN on the accuracy of global model

图 4 DBSCAN 中不同最小样本数对于全局模型准确率的影响

当最小样本数设置为 1 时, 实际上不存在离群点, 算法退化为基于 Sinkhorn 距离的层次化聚类, 此时模型性能最差. 而当最小样本数为 8 时, 由于聚类簇较少, 产生了较多的离群点. 而过多离群点被剔除同样导致模型性能不佳.

表 5 展示了 HS-CFA 对于 Sinkhorn 距离计算分布相似性的采样数目的影响. 实验使用的数据集为 CIFAR-10, 采样方法为随机采样, 采样样本数选择 4, 16, 64, 128 时对全局模型的准确率和相似度计算所需时间的影响. Sinkhorn 距离计算时间复杂度取决于采样数据组成成本矩阵的大小n和迭代次数k, 时间复杂度为$O(kn^2)$, 而经典最优传输代价 Wasserstein 距离的复杂度为$O(n^3\log n)$.

从表 5 中可以看到, 尽管采用了 GPU 加速, 但是

Table 5 Impact of Sampling Amounts on the Module Processing Time of Similarity Computation and Model Accuracy

表 5 采样数目对相似度计算模块处理时间和模型准确率影响

采样数目	相似度计算处理时间/s	模型准确率/%
4	0.278 31	70.96
16	1.133 26	74.87
64	3.245 51	75.43
128	8.798 12	75.51

随着采样数目的增加, 相似度计算的处理时间显著增长, 并且模型准确率在初期阶段有显著提升, 然而在采样数达到一定阈值后增加采样数对模型准确率的提升效果趋于饱和. 因此, 在实际应用中, HS-CFA 需要在计算开销与模型准确率提升之间权衡, 以选择合适的采样数目, 从而兼顾计算效率和模型性能.

5 结 论

针对联邦学习中 non-IID 导致的数据异构性问题, 本文提出了一种 2 层聚类联邦学习算法. 该算法采用轻量化的 Sinkhorn 距离来衡量不同客户端的数据分布相似性, 并在聚类过程中使用 DBSCAN 算法, 通过对簇内模型和簇间模型平均聚合优化全局模型. 该算法可有效地将具有相似数据分布的客户端进行自适应分簇进而动态修改本地模型聚合时的权重, 提高了联邦学习全局模型准确性和泛化能力. 实验结果表明, 本文提出的算法在处理数据异构性方面, 相较于传统联邦学习算法和其他基于聚类的算法在模型准确率和鲁棒性方面均取得了良好的效果.

总体而言本文通过结合最优传输代价和聚类算法, 为联邦学习数据异构性问题提供了了一种轻量且高效的优化方案. 未来的工作将重点关注如何设计更轻量级且定义良好的数据分布相似性度量方法, 以进一步降低计算开销并提升模型在边缘计算中的适应性.

作者贡献声明: 赵瑞提出方案整体设计、算法思路并撰写论文; 段小文负责实验验证; 刘新负责工程化指导; 周睿和周庆国指导论文并提出修改意见.

参 考 文 献

[1] Noor M M, Hassan W H. Current research on Internet of things (IoT)

security: A survey[J]. Computer Networks, 2019, 148: 283−294

[2] Kumar S R, Nystrom M, Lambert J, et al. Adversarial machine learning-industry perspectives[C]//Proc of the 41st IEEE Symp on Security and Privacy. Piscataway, NJ: IEEE, 2020: 69−75

[3] Sarkar S. Internet of Things[M]. San Francisco: Morgan Kaufmann, 2016

[4] Mao Yuyi, You Changsheng, Zhang Jun, et al. A survey on mobile edge computing: The communication perspective[J]. IEEE Communications Surveys & Tutorials, 2017, 19(4): 2322−2358

[5] Yu Wei, Liang Fan, He Xiaofei, et al. A survey on the edge computing for the Internet of Things[J]. IEEE Access, 2017, 6: 6900−6919

[6] Shakarami A, Ghobaei-Arani M, Shahidinegad A. A survey on the computation offloading approaches in mobile edge computing: A machine learning-based perspective[J]. Computer Networks, 2020, 182: 107496

[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] Abreha H G, Hayajneh M, Serhani M A. Federated learning in edge computing: A systematic survey[J]. Sensors, 2022, 22(2): 450

[9] Zhang Xueqing, Liu Yanwei, Liu Jinxia, et al. An overview of federated learning in edge intelligence[J]. Journal of Computer Research and Development, 2023, 60(6): 1276−1295 (in Chinese)
(张雪晴, 刘延伟, 刘金霞, 等. 面向边缘智能的联邦学习综述[J]. 计算机研究与发展, 2023, 60(6): 1276−1295)

[10] Zhu Hangyu, Xu Jinjin, Liu Shiqing, et al. Federated learning on non-IID data: A survey[J]. Neurocomputing, 2021, 465: 371−390

[11] Sattler F, Wiedemann S, Muller K R, et al. Robust and communication-efficient federated learning from non-IID data[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(9): 3400−3413

[12] Qu Liangqiong, Zhou Yuyin, Liang P P, et al. Rethinking architecture design for tackling data heterogeneity in federated learning[C]//Proc of the 2022 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2022: 10061−10071

[13] Zhang Ruilin, Du Jinhua, Yin Hao. Client selection algorithm in cross-device federated learning[J]. Journal of Software, 2024, 35(12): 5725−5740 (in Chinese)
(张瑞麟, 杜晋华, 尹浩. 跨设备联邦学习中的客户端选择算法[J]. 软件学报, 2024, 35(12): 5725−5740)

[14] Guo Songyue, Wang Yangqian, Bai Siyuan, et al. Federated adaptive interaction model for mixed distribution data[J]. Journal of Computer Research and Development, 2023, 60(6): 1346−1357 (in Chinese)
(郭松岳, 王阳谦, 柏思远, 等. 面向数据混合分布的联邦自适应交互模型[J]. 计算机研究与发展, 2023, 60(6): 1346−1357)

[15] Wang Jianyu, Charles Z, Xu Zheng, et al. A field guide to federated optimization[J]. arXiv preprint, arXiv: 2107.06917, 2021

[16] Ghosh A, Chung Jichan, Yin Dong, et al. An efficient framework for clustered federated learning[C]//Proc of the 33rd Advances in Neural Information Processing Systems. New York: Curran Associates Inc, 2020: 19586−19597

[17] Briggs C, Fan Zhong, Andras P. Federated learning with hierarchical clustering of local updates to improve training on non-IID data[C]// Proc of the 2020 Int Joint Conf on Neural Networks. Piscataway, NJ: IEEE, 2020: 1−9

[18] Kim Y, Hakim E A, Haraldson J, et al. Dynamic clustering in federated learning[C]//Proc of the 2021 IEEE Int Conf on Communications. Piscataway, NJ: IEEE, 2021: 1−6

[19] Li Chengxi, Li Gang, Varshney P K. Federated learning with soft clustering[J]. IEEE Internet of Things Journal, 2021, 9(10): 7773−7782

[20] Sattler F, Muller K R, Samek W. Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 32(8): 3710−3722

[21] Morafah M, Vahidian S, Wang Weijia, et al. FLIS: Clustered federated learning via inference similarity for non-IID data distribution[J]. IEEE Open Journal of the Computer Society, 2023, 4: 109−120

[22] Cuturi M. Sinkhorn distances: Lightspeed computation of optimal transport[C]//Proc of the 26th Advances in Neural Information Processing Systems. New York: Curran Associates Inc, 2013: 2292−2300

[23] Schubert E, Sander J, Ester M, et al. DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN[J]. ACM Transactions on Database Systems, 2017, 42(3): 1−21

[24] Sejourbe T, Feydy J, Vialard F X, et al. Sinkhorn divergences for unbalanced optimal transport[J]. arXiv preprint, arXiv: 1910.12958, 2019

[25] Mcmahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Proc of the 20th Int Conf on Artificial Intelligence and Statistics. New York: PMLR, 2017: 1273−1282

[26] Wan Wei, Hu Shengshan, Lu Jianrong, et al. Enhancing generalization robustness of federated learning in highly heterogeneous environments[J]. SCIENTIA SINICA Informationis, 2024, 54(3): 566−581 (in Chinese)
(万伟, 胡胜山, 陆建荣, 等. 联邦学习在高度数据异构场景下的泛化鲁棒性增强[J]. 中国科学: 信息科学, 2024, 54(3): 566−581)

[27] Li Tian, Sahu A K, Zaheer M, et al. Federated optimization in heterogeneous networks[J]. Proceedings of Machine Learning and Systems, 2020, 2: 429−450

[28] Karimireddy S P, Kale S, Mohri M, et al. SCAFFOLD: Stochastic controlled averaging for federated learning[C]//Proc of the 37th Int Conf on Machine Learning. New York: PMLR, 2020: 5132−5143

[29] Yan Bingjie, Liu Boyi, Wang Lujia, et al. FedCM: A real-time contribution measurement method for participants in federated learning[C]//Proc of the 2021 Int Joint Conf on Neural Networks. Piscataway, NJ: IEEE, 2021: 1−8

[30] Sun Yan, Shen Li, Huang Tiansheng, et al. FedSpeed: Larger local interval, less communication round, and higher generalization accuracy[C/OL]//Proc of the 11th Int Conf on Learning Representations. 2023[2025-03-03]. https://openreview.net/forum?id=bZjxxYURKT

[31] Ghadimi E, Teixeira A, Shames I, et al. Optimal parameter selection for the alternating direction method of multipliers (ADMM): Quadratic problems[J]. IEEE Transactions on Automatic Control, 2014, 60(3): 644−658

[32] Fama F, Kalalas C, Lagen S, et al. Measuring data similarity for

efficient federated learning: A feasibility study[C]//Proc of the 13th Int Conf on Computing, Networking and Communications. Piscataway, NJ: IEEE, 2024: 394−400

[33] Xiao Han, Rasul K, Vollgraf R. Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms[J]. arXiv preprint, arXiv: 1708.07747, 2017

[34] Krizhevisky A, Hinton G E. Learning multiple layers of features from tiny images[R]. Toronto: University of Toronto, 2009

[35] Hsu T H, Qi Hang, Brown M. Measuring the effects of non-identical data distribution for federated visual classification[J]. arXiv preprint, arXiv: 1909.06335, 2019

[36] Krizhevisky A, Sutskever I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84−90

[37] Paszke A, Gross S, Massa F, et al. PyTorch: An imperative style, high-performance deep learning library[C]//Proc of the 33rd Int Conf on Neural Information Processing Systems. New York: Curran Associates Inc, 2019: 8026−8039

[38] Reisizadeh A, Farnia F, Pedarsani R, et al. Robust federated learning: The case of affine distribution shifts[C]//Proc of the 34th Int Conf on Neural Information Processing Systems. New York: Curran Associates Inc, 2020: 21554−21565



Zhao Rui, born in 1989. PhD candidate. Student member of CCF. His main research interests include distributed neural networks and vehicular-roadside cooperation.

赵 瑞, 1989 年生. 博士研究生. CCF 学生会员. 主要研究方向为分布式神经网络、车路协同.



Duan Xiaowen, born in 2000. Master candidate. Her main research interest includes federated learning algorithm optimization.

段小文, 2000 年生. 硕士研究生. 主要研究方向为联邦学习算法优化.



Liu Xin, born in 1995. PhD, associate professor. Member of CCF. His main research interests include cybersecurity and artificial intelligence.

刘 新, 1995 年生. 博士, 副教授. CCF 会员. 主要研究方向为网络安全、人工智能.



Zhou Rui, born in 1981. PhD, professor. Senior member of CCF. His main research interests include distributed real-time systems and artificial intelligence.

周 睿, 1981 年生. 博士, 教授. CCF 高级会员. 主要研究方向为分布式实时系统、人工智能.



Zhou Qingguo, born in 1973. PhD, professor. Senior member of CCF. His main research interests include distributed and embedded real-time systems, neural networks, and intelligent transportation.

周庆国, 1973 年生. 博士, 教授. CCF 高级会员. 主要研究方向为分布式与嵌入式实时系统、神经网络、智慧交通.