计算机学报
CHINESE JOURNAL OF COMPUTERS

基于可学习聚合权重的解析性联邦学习方法

蒋伟进[1)2)] 崔新雨[1)2)] 刘志华[1)2)] 陈伸有[1)2)] 胡佳龙[1)2)]

[1)](湖南工商大学计算机学院, 长沙 410205)

[2)](湘江实验室, 长沙 410205)

摘要 联邦学习通过在客户端与参数服务器之间交换模型参数而非原始数据，有效保护了数据隐私安全。然而，随着客户端数量和数据规模的增加，联邦学习仍面临通信开销增加和任务复杂性提升的问题。现有方法通常采用基于客户端本地数据量的权重归一化策略进行模型聚合，在一定程度上降低通信开销，但未充分考虑数据异质性，这可能导致模型过拟合、收敛速度减缓，并加重通信负担。因此，本文提出了一种具有可学习聚合权重的解析性联邦学习算法（Learnable Aggregation Weights and Analytic Federated Learning，LAW-AFL），该算法首先通过引入可学习的收缩因子和相对权重，改进了聚合过程中的权重计算方式，并引入闭式训练范式指导神经网络训练，增强模型在异质性数据下的稳定性和泛化能力；其次通过推导绝对聚合规则，进一步提升了聚合过程的效率和准确性，实现了单周期本地训练，简化了训练流程，同时该算法利用闭式解进行高效聚合，简化了训练流程。实验结果表明，所提出的算法在多个数据集和模型上都显著提高了全局模型的精度和泛化能力，相比较于基线方法，在处理大规模客户端和非独立同分布（Non Independent and Identically Distributed，Non-IID）数据时准确率提高了 10%，并在特定实验设置下将全局模型的准确率提升至 90%以上，单论训练时间相较于 FedAVG 缩短了 69.82 秒/轮。这证明了 LAW-AFL 在准确性和鲁棒性方面具有一定的优势，并且大幅度降低了通信成本。

关键词 联邦学习；可学习聚合权重；闭式训练范式；自动解析技术；泛化能力；通信成本

中图法分类号 TP18

Analytical Federated Learning Method Based on Learnable Aggregation Weights

JIANG wei-jin[1)2)] CUI xin-yu[1)2)] LIU zhi-hua[1)2)] CHEN shen-you[1)2)] HU jia-long[1)2)]

[1)]（Department of Computer Science, Hunan University of Technology and Business, Changsha 410205）

[1)](Xiangjiang Laboratory, Changsha 410205)

Abstract Federated learning protects data privacy by exchanging model parameters rather than raw data between clients and a central server. However, as the number of clients and the volume of data grow, it still faces increasing communication overhead and task complexity. Existing methods typically normalize aggregation weights based on each client's local data size to reduce communication cost, but they often overlook data heterogeneity, which can lead to overfitting, slower convergence, and greater overall communication burden.To address these issues, we propose Learnable Aggregation Weights and Analytic Federated Learning (LAW-AFL). First, LAW-AFL introduces a learnable shrinkage factor and relative weights to refine the aggregation process, and employs a closed-form training paradigm to guide neural network optimization, thereby enhancing model stability and generalization under heterogeneous data. Second, by deriving an absolute aggregation rule, it further improves aggregation efficiency and accuracy, enables single-pass local training, and simplifies the overall training pipeline through closed-form updates.Extensive experiments on multiple datasets and model architectures show that LAW-AFL significantly improves global model accuracy and generalization. On large-scale, non-IID data, it achieves a 10% increase in accuracy

---

本课题得到国家自然科学基金(61772196)；湖南省自然科学基金(2020JJ4249)，湖南省教育厅科学研究重点项目(24A0446;24A0753)，长沙市社科联哲学社会科学规划课题(2024CSSKKT31)资助. **蒋伟进**，博士，教授，博士生导师，中国计算机学会（CCF）高级会员，博士生导师，主要研究领域为边缘计算、联邦学习、网络空间安全。**崔新雨**(通信作者)，硕士研究生，CCF 学生会员，主要研究领域为联邦学习、隐私保护。**刘志华**，硕士研究生，CCF 学生会员，主要研究领域为联邦学习、隐私保护。**陈伸有**，硕士研究生，CCF 学生会员，主要研究领域为联邦学习、隐私保护。**胡佳龙**，硕士研究生，CCF 学生会员，主要研究领域为数据动态定价、联邦学习。

compared to existing methods and exceeds 90% accuracy under specific experimental settings, while reducing per-round training time by 69.82 seconds relative to FedAVG. These results demonstrate that LAW-AFL offers clear advantages in accuracy, robustness, and communication efficiency.

Key words Federated Learning; Learnable Aggregation Weights; Closed-form Training Paradigm; Automated Analytical Techniques; Generalization Capability; Communication Cost

1 引言

联邦学习（Federated Learning,FL）[1],[2]作为一种去中心化的机器学习框架，因其优越的数据隐私保护特性而广泛应用。相较于传统的集中式训练，联邦学习将模型的训练过程分布在多个设备上。各客户端利用本地数据进行模型训练，仅将模型更新上传至服务器进行聚合以更新全局模型。这种分布式设计在一定程度上避免了传统集中式学习框架下可能出现的数据泄露风险。联邦学习的引入有效解决了数据隐私保护和数据中心化的问题，同时能够利用多方数据，提高了模型的泛化性和准确率[3]。联邦学习通过融合本地计算与模型传输的理念，规避了传统中心化机器学习框架下潜在的隐私信息泄露风险[4]。

然而，联邦学习在实际应用中仍面临诸多挑战，尤其是在数据呈现非独立同分布（Non Independent and Identically Distributed，Non-IID）和大规模客户端参与的场景下，其性能和通信效率往往受到很大影响。现有方法大多基于固定的模型聚合策略，FedAVG[2]作为 FL 中最经典的聚合算法之一，采用基于客户端本地数据样本数量的加权平均策略：服务端先把统一模型架构发给各客户端，客户端利用本地数据训练并上传更新，服务端按各客户端数据样本数量加权平均这些参数生成全局共享模型[5]。上述算法假设客户端的数据质量和分布均等，但在实际场景中可能存在数据分布不均或质量差异等异质性问题，因此这种固定聚合策略未能充分考虑到不同本地模型因数据分布不均而带来的贡献度差异，导致全局模型的泛化能力下降。

为应对数据异质性问题，Kong 等人[6]提出一种基于知识蒸馏的联邦学习框架，该框架的服务器端通过全局蒸馏整合客户端的预测分布，有效缓解了 Non-IID 数据对模型性能的影响，但其对公共数据集的依赖以及蒸馏训练也增加了额外通信开销。为了降低通信成本，Lin 等人[7]通过将边缘计算和联邦学习的结合，降低能耗并减少通信延迟，但其方法未深入解决模型更新本身可能存在的隐私泄露风险，导致攻击者可以通过分析上传的本地模型推断出参与者的训练数据 **Error! Reference source not found.**，而聚合服务器也能够从聚合后的全局模型中提取参与者上传数据的统计特征[9]，导致数据隐私泄露的风险。因此，为确保数据隐私安全，必须采取有效的隐私保护技术对模型参数进行加密和保护。

现有若干关于隐私保护的联邦学习方案[10-12]在保护隐私方面取得了一定成效，但往往难以同时适应异构数据环境并在模型性能、计算开销与通信成本之间达到良好平衡。例如：为了节省资源，聚合服务器可能只对部分梯度进行聚合[13]，这将导致模型性能下降或训练无法收敛。

基于对联邦学习系统的研究分析，本文得出以下两点发现：

（1）固定的聚合策略无法充分反映客户端的实际贡献，而通过引入可学习的聚合权重，可以动态调整每个客户端的权重，提升全局模型的泛化能力和稳定性，从而更好地应对数据异质性和大规模客户端的挑战。

（2）现有方法在应对模型异构性和数据异质性时，面临着模型精度下降的严峻挑战，需要进一步的优化和改进以提升模型精度和鲁棒性。

基于上述发现，本文提出了具有可学习聚合权重的解析性联邦学习算法（Learnable Aggregation Weights and Analytic Federated Learning，LAW-AFL），将自动加权聚合与解析性训练有机结合，有效解决模型因数据分布差异带来的高通信成本的挑战；简化了训练过程，缩短了训练时间，提高了模型的鲁棒性和准确度。在多种数据集和模型架构的对比中，LAW-AFL 通过学习最优聚合权重增强了全局模型在异质数

据下的泛化性能；解析性单轮聚合进一步压缩了训练时间并提高了计算效率。面对数据分布差异加剧和客户端规模扩大，LAW-AFL 在准确率和波动控制方面均优于各基线方法。此外，LAW-AFL 在极端 Non-IID 数据分布和大量客户端的情况下表现稳定，展示了其在复杂数据环境下的鲁棒性以及在通信效率、模型性能和隐私保护方面的优势。

本文的主要贡献如下：

（1)提出了一种具有可学习聚合权重的解析性联邦学习算法，该算法将自动加权聚合与解析性训练有机结合，使其能够在单轮聚合中实现全局模型更新。这种结合不仅提高了模型的鲁棒性和准确度，还降低了通信成本。

（2)通过学习最优聚合权重优化全局模型泛化性能，并利用解析性训练缩短训练时间；在极端 Non-IID 数据分布及大量客户端场景下依旧保持稳定表现。

（3)引入差异化的噪声添加策略与差分隐私保护方法，在不牺牲模型性能的前提下，提供了更好的隐私保障；这一策略有效地隐藏了客户端的梯度信息，防范潜在的数据泄露。

2 相关工作

2.1 联邦学习

FL是一种允许模型分散在许多移动设备上进行训练的分布式优化范例，通过聚合单个训练权重来集体训练数据孤岛上的机器学习模型，同时还保留源数据的隐私性。在FL中，每个客户端可以在本地保存其数据集，仅定期共享其本地训练的模型更新，中央服务器聚合客户端的局部梯度进行协同训练，然后对局部模型进行加权聚合生成全局模型。在FedAVG提出之后，为解决联邦学习中因Non-IID数据导致的性能下降问题，各种改进方法相继被提出。例如：FedPROX[14]通过引入正则化项限制本地更新的幅度，以减少本地数据偏差对全局模型的影响；而FedDYN[15]通过动态正则化调节Non-IID数据对模型优化的影响，不同于FedPROX的静态正则化或FedDF的固定蒸馏权重，现有隐私保护方法（如：DP-FedAVG[12]）采用均匀噪声。然而这些研究在数据异质性和系统异构性还存在一定的局限。另一类方法则侧重于设计自适应聚合权重，以优化从多个客户端获取的模型融合效果。

文献[16]提出的重新审视使用神经网络进行联邦学习中的加权聚合（Revisiting Weighted Aggregation in Federated Learning with Neural Networks，FedLAW）是一种针对Non-IID数据的动态加权聚合算法。其核心思想是通过客户端本地模型的性能表现动态调整聚合权重，从而优化全局模型的泛化能力。在数据异质性情况下，优化聚合权重。提高模型泛化能力。然而FedLAW依赖静态收缩因子和梯度优化权重，导致计算效率低下且缺乏隐私保护，并且客户端的数据分布与全局数据分布不一致，使得客户端模型在全局测试集上的表现不如本地数据的表现。随着数据量的增大，模型训练时间和计算成本显著增加。

2.2 分析学习

分析学习（Analytic Learning，AL）作为基于梯度更新相关问题的策略，常用于解决传统梯度下降训练过程中的迭代收敛和训练时间问题。分析学习通过矩阵伪逆（pseudoinverse）直接求解全局模型参数，它也被称为伪逆学习[17]。分析学习的起点是浅层学习，径向基网络（Radial Basis Network，RBFN）在第一层执行核变换后，使用正交最小二乘[18]估计（Least Squares，LS）来训练参数。在深度网络中，多层分析学习[19][20]引入单周期训练方式，利用最小二乘技术逐层训练堆叠的自编码器，如密集伪逆自编码器（Dense Pseudoinverse Autoencoder，DPA）[21]，它使用最小二乘解来结合浅层和深层特征，以逐层训练堆叠的自编码器。

然而，早期的分析学习技术在训练权重时需要同时处理整个数据集，因此面临内存挑战。这种内存问题通过分块递归的Moore-Penrose逆[22]（Block-wise Recursive Moore-Penrose Inverse，MP）得到缓解，这种递归等效特性与持续学习需求相呼应，分析学习还被应用于持续学习，利用其等效特性处理序列数据分割和灾难性遗忘问题。本文借鉴这些适应性特性，尝试将类似的解析性策略引入联邦学习，以提高不同客户端数据分布下的模型性能一致性。

2.3 解析性联邦学习

文献[23]提出了可解析性联邦学习（Analytic Federated Learning，AFL），AFL是一种基于闭式解（closed-form solution）的联邦学习范式，其核

心思想是通过闭式解直接求解全局模型参数，避免传统梯度下降的迭代过程。与传统联邦学习方法（如FedAVG）相比，AFL具有单周期训练、无超参数、理论收敛保证的优势；AFL也是一种无需梯度的FL框架，具有分析解决方案，减少通信成本和训练次数。然而AFL的闭式聚合策略难以适应Non-IID数据分布，并未阻止模型的过拟合现象，无法根据客户端动态贡献调整聚合权重，限制了模型优化的灵活性。为此，本文提出可学习聚合权重机制嵌入AFL框架通过动态调整客户端权重来缓解数据异质性影响，能够根据客户端数据分布差异自动分配权重，通过权重收缩因子γ控制模型复杂度，缓解过拟合。并且差异化噪声策略与动态权重协同，在不增加通信开销的前提下提升隐私保护。

当前，大部分研究聚焦于将聚合权重归一化，并与局部数据大小成比例，然而由于非凸性[24]，过度参数化[25]以及深度神经网络（Deep neural networks，DNN）的独特性质，全局权重收缩在每轮的学习中设置一个动态收缩因子γ来收缩全局模型的参数。文献 [26]基于DNN的输入尺度不变性验证了该方法在理想条件下的有效性。但受限于复杂数据分布和有限样本，本文提出了针对DNN训练过程的改进策略，旨在提升模型在多种任务和数据集上的泛化能力。调整聚合权重γ的值，在正则化和优化之间找到平衡，从而保持模型性能的同时提高隐私保护。

关于联邦学习的模型聚合方面。文献[27]提出了AUTO-FedAVG通过学习不同的机构医疗数据的聚合权重来实现个性化医疗，文献[28]通过学习局部数据集的聚合权重来匹配分散FL中的相似对等体。现有研究工作普遍设定$\gamma = 1$实现归一化聚合，而本文通过引入自适应γ，提升模型在异质数据条件下的整体泛化表现。除此之外，FL假设在每个客户端中存在多个本地训练周期，并且客户端通常具有异构数据，在这种情况下，客户端的局部梯度在Non-IID数据下呈现低相干性[1]，他们无法从学习权重中理解FL的动态，为进一步研究联邦学习动态提供方向。因此本研究采用客户端数据局部处理机制，保证数据在客户端进行更多的处理和分析，动态的调整隐私保护策略，在数据分布不均匀的情况下增加噪声水平，保护数据的隐私安全。

本文中采用DNN等神经网络作为模型架构，在探讨FL框架下的神经网络训练动态时，全局权重缩减策略与客户端模型一致性机制，其作用机制与集中式学习中的权重衰减和梯度同步策略具有相似性。权重衰减是一种正则化技术，有助于防止模型在训练数据上过拟合，进而提升模型在新数据上的泛化能力。实验结果表明，融合FedLAW与AFL框架，最优权重缩减因子与训练周期数量近似成反比。这意味着，当训练过程跨越更多的周期时，实施权重缩减的必要性会有所减弱。通过将权重缩减与FL的聚合机制相结合，不仅能够控制全局模型的复杂度，还能在一定程度上促进不同客户端模型之间的参数一致性。

3 可学习聚合权重的自适应解析性联邦学习

3.1 自适应全局收缩与训练动态

3.1.1 定义

在 FL 中，客户集合用 S 表示，客户 i 的局部数据集用 $\mathrm{D}_i = \{(x_j, y_j)\}_{j=1}^{N_i}$ 表示，其中 N_i 为客户端 i 的数据样本数量。整个系统的数据集 D 为所有客户端数据的并集：$D = \cup_{i \in S} D_i$。θ为聚合权重。

本文使用的主要符号说明如表 1：

表 1 主要符号说明

符号	释义
w_g^t	第 t 轮通信后的全局模型参数
η_g	全局学习率
η_l	局部学习率
γ	收缩因子
S	参与训练的客户端集合
λ_i	第 i 个客户端的相对权重
w_i^t	客户端i在第t轮的本地模型参数
$w_{agg,k}$	累积聚合到第i个客户端的中间权重矩阵
n_k	第i个客户端的数据量
V	参与聚合的客户端集合
n	所有客户端数据总量
P	每个客户端的类别分布
α	Dirichlet分布的参数
$X^\dagger$	矩阵X的伪逆矩阵
X	分块矩阵
X_u，X_v	矩阵X的上下块部分
W_u，W_v	权重分量，由X_u和X_v计算得到

$\widehat{W}_u$，$\widehat{W}_v$	聚合时的加权系数，决定W_u和W_v对整体W的贡献
$L_{proxy}(\{\lambda,\gamma\})$	客户端模型加权后的预测误差

（1）全局模型聚合规则：设客户端集合 S 的本地模型参数为$\{w_i^t\}$，收缩因子$\gamma>0$，且权重向量λ满足$\lambda_i \geq 0$ 和$\|\lambda\|_1=1$，则全局模型更新为

$$w_g^t=\gamma\sum_{i=1}^{m}\lambda_i\, w_i^t \#(1)$$

在每一轮中，客户的局部模型都被初始化为$w_i^t \leftarrow w_g^t$的全局模型，客户并行进行局部训练。在每个局部训练时期，客户端以局部学习率η_l进行随机梯度下降（Stochastic Gradient Descent，SGD）更新，每次 SGD 迭代显示为

$$w_i^t \leftarrow w_i^t-\eta_l\nabla\ell(B_k,w_i^t),\quad for\ k=1,2,\dots,K \#(2)$$

其中ℓ是损失函数，B_k是第 i 次迭代时从D_i采样的小批量。客户端本地更新后，服务器对 m 个客户端进行采样聚合，客户端 i 的本地更新伪梯度表示为$g_i^t=w_g^t-w_i^t$。然后服务器进行加权聚合，将局部模型合并更新为新的全局模型如公式（3）

$$w_g^{t+1}=\sum_{i=1}^{m}\mu_i\, w_i^t=\|\mu\|_1 w_g^t-\eta_g\sum_{i=1}^{m}\mu_i\, g_i^t \quad s.t.\ \mu_i\geq 0, \#(3)$$

其中，$\mu=[\mu_1,\dots,\mu_m]$是权重向量，$\|\mu\|_1$是L_1范数；即各元素绝对值之和，用于约束聚合权重归一化，确保全局模型更新的稳定性，$\eta_g=1$是全局学习率。对于基本的 FedAVG，它采用与数据规模成比例的归一化权重，即$\mu_i=\frac{|D_i|}{|D|}$，$D=\cup_{i\in S}D_i$，在本文中，假设权重向量未进行归一化（$\|\mu\|_1\neq 1$），这意味着l_1范数不一定等于 1。通过将μ分解为$\{\gamma,\lambda\}$来独立研究l_1范数和相对权重的影响。

因此，公式（3）重新表述为

$$w_g^{t+1}=\gamma\sum_{i=1}^{m}\lambda_i\, w_i^t \quad s.t.\ \gamma>0,\lambda_i\geq 0 \#(4)$$

当γ<1 时，这将导致全局模型的权重收缩，因此在这种情况下，称γ为收缩因子。

（2）收缩因子γ的自适应规则：设客户端梯度为$g_i=\nabla L_i(w_g)$，并定义其平方范数的期望为$E[\|g_i\|^2]$，收缩因子的定义为

$$\gamma=\frac{\eta_g}{E[\|g_i\|^2]+\epsilon}\#(5)$$

其中η_g是全局学习率，ϵ>0 是用于平滑的常数，保证分母不为零。

由公式（5）可知，γ与$E[\|g_i\|^2]$成反比，当$E[\|g_i\|^2]$较大时，即梯度幅度大，表示在优化曲线上局部斜率较陡，分母增大，使得γ较小，这意味着在陡峭区域采用较小步长，避免因更新步长过大而导致发散。反之当$E[\|g_i\|^2]$较小时，梯度幅度小，接近局部最优或者平坦区域，使得γ较大，有利于加速收敛。

在实际的 FL 场景中，客户端之间的数据分布通常为 Non-IID，即每个客户端的数据分布存在差异。客户端间的数据往往呈现出 Non-IID 的特性[30]。数据的 Non-IID 特性对 FL 的收敛速度构成了严峻挑战。

定义 1（Dirichlet 分布与客户端异质性）：本文采用 Dirichlet 抽样[31]来模拟客户异质性。Dirichlet 抽样被广泛用于模拟客户端之间的数据异质性。这种抽样方法通过调整 Dirichlet 参数α来控制数据分布的异质性程度。其概率密度函数为：

$$p(\boldsymbol{p};\alpha)=\frac{1}{B(\alpha)}\prod_{k=1}^{K}{p_k}^{\alpha_k-1}\#(6)$$

（3）加权聚合更新推导：

将全局模型的更新归纳为两部分：一是“伪梯度”收缩项$(1-\gamma)w_g^t$，二是加权平均梯度项$\gamma\sum_{i=1}^{m}\lambda_i\, w_i^t$。

$$w_g^{t+1}=\gamma\sum_{i=1}^{m}\lambda_i\, w_i^t\#(7)$$

其中$\lambda_i=\frac{|D_i|}{|D|}$,$\sum_{i=1}^{m}\lambda_i=1$表示第 i 各客户端的相对数据量占比。

最佳收缩因子：

在服务器端利用代理数据集 P 构造目标损失

$$\mathcal{L}_{proxy}\left(\gamma\sum_{i=1}^{m}\lambda_i\, w_i^t\right)\#(8)$$

并通过

$$\gamma^*=\arg\min_{\gamma>0}\mathcal{L}_{proxy}\left(\gamma\cdot\sum_{i=1}^{m}\lambda_i w_i^t\right)\#(9)$$

求得最优γ^*，该值平衡了“全局梯度优化”

与“模型权重正则化”两者作用。

推导与参数说明

（1）传统全局更新到收缩：

$$w_g^{t+1} = w_g^t - \eta_g g_g^t \#(10)$$

令γ对更新整体收缩，可写为

$$w_g^{t+1} = \gamma\left(w_g^t - \eta_g g_g^t\right) = w_g^t - \gamma\eta_g g_g^t - (1-\gamma)w_g^t \#(11)$$

将$(1-\gamma)w_g^t$称为全局权重收缩的伪梯度部分，$\gamma\eta_g g_g^t$则为全局平均梯度。

（2）相对权重λ_i

采用加权平均代替梯度g_g^t：

$$w_g^t - \eta_g g_g^t \to \sum_{i=1}^{m} \lambda_i w_i^t, \lambda_i = \frac{|D_i|}{|D|} \#(12)$$

(3)收缩因子优化

理论上可用梯度下降法在代理数据集上求解公式（9），实际中因 P 规模受限，改用启发式预实验确定经验最优γ，既保证计算效率，也兼顾性能。

3.1.2 算法设计与实现

本文从自动加权聚合的角度来研究对 FL 训练的影响，旨在优化聚合过程中的权重分配。自动加权聚合算法的核心在于为每个客户端的模型分配一个动态的权重，这个权重是基于该客户端的数据质量、模型性能以及其他相关因素来确定的。通过数据驱动的方式动态调整权重，可以实现更高效的模型聚合，从而提高全局模型的性能。引入相对权重来调整每个客户端的贡献。混合更新策略是指对于被选中的客户端，使用它们更新后的模型参数来更新全局模型；对于未被选中的客户端，则保持它们上一轮的模型参数不变，并按其数据量比例贡献到全局模型中。公式如下：

$$w_g^{t+1} = \left(\sum_{k\notin V} \frac{n_k}{n} w_g^t\right) + \left(\sum_{k\notin V} \frac{n_k}{n} w_g^{t+1}\right) (k = 1,2,\dots,m) \#(13)$$

不同于传统的联邦平均算法采用固定且归一化的权重，本文采用的聚合权重μ自适应地调整，用以精确反映各客户端数据的质量和对全局模型的贡献度。为了区分解耦权重的总体大小（即l_1范数）和相对权重（即各权重在总权重中的比例），本文采用一种分解策略，将μ分解为两个组成部分：γ和λ。其中，$\gamma = \|\mu\|$表示收缩因子，而$\lambda = [\lambda_1, \dots, \lambda_m]$满足$\lambda_i = \frac{\mu_i}{\gamma}$，表示各客户端的相对权重。通过这样的分解，可以独立地研究收缩因子γ和相对权重向量λ对全局模型训练过程的影响。在加权聚合策略的框架下，全局模型的更新可以表示为以下公式：

$$w_g^t = w_g^{t-1} + \eta_g \cdot \gamma \cdot \sum_{i\in S} \lambda_i \cdot w_i^t \#(14)$$

在自动加权聚合策略中，本文引入了一个额外的参数γ（收缩因子），用于动态地调整全局模型的更新幅度。参数γ的取值范围限定在(0,1]区间内，确保了模型更新的适度性，避免了过大或过小的更新步长对模型稳定性和收敛性的影响。最优的收缩因子γ和相对权重λ通过梯度下降法在代理数据集上进行学习。通过调整γ，本文可以在正则化和优化之间找到一个平衡，从而改善全局模型的性能。

3.1.3 实现步骤

自动加权聚合算法的具体实现步骤包括初始化，权重调整，模型更新等环节。下面将会详细的解释和说明：

首先定义全局模型参数为w_g^t，设第 i 个客户端的本地模型参数为w_i^t。其次进行定义第 i 个客户端的初始权重为λ_i^1，为每个客户端的模型分配一个动态的权重，这个权重是基于该客户端模型在全局目标一致的代理数据集上的表现来确定的，在传统的自动加权聚合算法中，是利用梯度下降方法在代理数据集上学习这些权重，使得它们能够反映客户端模型对全局目标的贡献。本文通过闭式解替代梯度优化，降低了权重聚合的计算复杂度。最后采用加权平均的聚合规则，即：

$$w_g^t = \sum_{i-1}^{N} \lambda_i w_i^t \#(15)$$

在本文中客户端的初始权重，根据客户端的数量、数据质量以及模型性能等因素来进行初始化，其公式如下：

$$\lambda_t^{'} = \frac{q_i a_i}{\sum_{j=1}^{N} q_j a_j + \epsilon} (i = 1,2,\dots,N) \#(16)$$

本文使用客户端数量 N，客户端的数据质量为q_i，模型性能评估可以从两方面评估每个客户端的模型性能，一个是准确率a_i和损失值l_i。然后可以根据上述的因素计算每个客户端的权重。公式（15）是考虑了客户端数量 N、数据质量q_i以及准确率a_i三个因素来调整权重，可以加入损失值进一步考虑权重的调整，其公式如下：

$$\lambda_i = \frac{q_i a_i N \cdot \frac{1}{1+l_i}}{\sum_{j=1}^{N}(q_i a_i N \cdot \frac{1}{1+l_i})} \#(17)$$

每个客户端使用其本地数据训练模型，得到更新后的本地模型参数w_i^t，然后使用加权平均规则聚合更新后的本地模型参数，得到全局模型参数w_g^t，接着利用公式（6）规则进行聚合更新，其公式如下：

$$w_g^t = \sum_{i=1}^{N} w_i^t \#(18)$$

再将更新后的全局模型参数w_g^t分发回给每个客户端，使他们可以在下一次迭代中使用。在训练结束之后，输出最终的全局模型参数，作为训练的结果，然后在不同数据集上和不同神经网络模型架构上实验，得出最终结果。

算法 1 根据收缩因子γ计算自适应聚合权重μ_i，避免单个客户端对全局模型的过度影响。使用自适应聚合权重μ_i对客户端的局部模型参数w_i^t进行加权平均，更新全局模型参数$\widetilde{w}_g^t$，在加权聚合过程中，本文使用解析解进行聚合，避免了多次迭代，从而提高了聚合过程的效率和准确性。

自动加权聚合算法见算法 1。

算法1. 自动加权聚合

输入：w_g^t，w_i^t，γ

输出：更新后的全局模型参数$\widetilde{w}_g^t$

1：初始化全局模型参数w_g^t，初始聚合权重μ_i对所有i设置相同的默认值

2：FOR 每个客户端i DO

3：客户端i基于数据质量q_i和模型性能a_i计算相对权重λ'_i

$$\lambda'_i = \frac{q_i \cdot a_i}{\sum_{j=1}^{m} q_j \cdot a_i + \epsilon}$$

4：λ_i反映了客户端i的数据质量、数据量等因素

5：服务器对接收的权重进行标准化，确保$\sum_{i=1}^{N}\lambda_i = 1$:

$$\lambda_i = \frac{\lambda'_i N \frac{1}{1+l_i}}{\sum_{j=1}^{N}\lambda'_i N \frac{1}{1+l_i}}$$

6：根据收缩因子γ计算自适应聚合权重μ_i

7：$\mu_i = \gamma \cdot \lambda_i$

8：使用自适应权重μ_i进行全局模型参数w_g^t的加权聚合：

9：$\widetilde{w}_g^t \leftarrow \gamma^* \cdot \sum_{i=1}^{N} w_i^t \mu_i$

10：END FOR

3.2 解析性联邦学习

解析性联邦学习（Analytic Federated Learning, AFL）是一种全新的联邦学习训练范式，它融合了解析性学习的优势，为联邦学习领域带来了闭式（或称解析）解决方案。解析性联邦学习旨在通过聚合多个数据孤岛上的个体训练权重来共同训练一个机器学习模型，同时保护源数据的隐私。这一范式适用于对数据隐私保护要求极高的敏感领域，与传统联邦学习相比，解析性联邦学习的核心在于其引入了闭式解决方案进行网络训练，无需依赖梯度下降等迭代方法。AFL 的算法框架主要包括局部训练阶段和集中聚合阶段两个部分。接下来，将详细描述这两个阶段的算法过程。在局部训练阶段，w_i预训练主干作为特征提取器，促进人工智能网络学习，使训练在一个 Epoch 内完成。

3.2.1 局部训练阶段

局部训练阶段与传统联邦学习相似，在局部训练阶段，每个客户端首先在公开数据集上训练特征提取器，确保其具备通用特征表示能力；接着本文采用 ResNet-50 作为骨干网络，通过监督学习完成预训练，使用 SGD 优化器，设置初始学习率为 0.1，余弦衰减调度，训练 100 个周期。在联邦学习过程中，特征提取器的参数$\widetilde{w}_i$保持固定，仅对顶层L_i进行微调，以减少本地计算开销并避免过拟合。

在局部训练阶段采用预训练的骨干网络作为特征提取器，并通过微调优化，以加速训练过程。每个客户端初始化一个本地模型，该模型包括一个预训练的骨干网络和一个可训练的分类头。本文所有模型均采用相同的预训练模型作为特征提取器。固定预训练骨干网络的参数，仅对分类头（全连接层）进行训练，确保所有方法在相同的初始化条件下优化。在局部训练阶段引入动态特征提取器微调策略，AFL 使用数据集预训练骨干网络作为特征提取器。预训练的网络具备较强的特征提取能力，使用预训练的骨干网络对本地数据进行特征提取，这可以减少训练时间并提高模型性能。算法 2 描述解析性联邦学习局部训练的过程，将所有客户端局部模型$w_i(0)$和特征提取器$\widetilde{w}_i(0)$设置为相同的初始值，每个客户端在本地数据集D_i上独立训练模型。由于使用了预训练的特征提取器$\widetilde{w}_i$，AFL 只需对局部模型的上层参数w_i进行更新。其公式如下：

$$w_i{}^t = w_i{}^{t-1} - \eta \nabla_{w_i{}^t} L_i\left(w_i{}^t, \widetilde{w}_i{}^t\right) \#(19)$$

AFL的局部训练算法见算法2。

算法2. AFL的局部训练

输入：τ，T

输出：全局模型参数w_g^t

1：初始化w_g^t，并将$w_i(0)$和$\tilde{w}_i(0)$设为相同的初始值

2：FOR $t = 1,2,\dots,T\ DO$

3：　FOR *每个客户端* i *in parallel DO*

4：使用预训练的特征提取器$\tilde{w}_i$作为特征提取模块

5：根据客户端i的本地数据分布差异计算微调层数L_i（公式18）

6：　END FOR

7：在本地训练时，仅对顶层L_i层进行微调

8:在局部数据集D_i上进行局部训练，更新局部模型参数${w_i}^t$：

9：${w_i}^t = {w_i}^{t-1} - \eta\nabla_{{w_i}^t} L_i\left({w_i}^t, {\tilde{w}_i}^t\right)$

10：END FOR

如算法 2 所示，其中L_i是在客户端 i 上的损失函数，η是学习率。通过只微调最后几层的参数，AFL 减少了训练时间和计算资源消耗。

3.2.2　集中聚合阶段

集中聚合阶段是 AFL 的核心之一。传统的联邦学习通常使用加权平均（如 FedAVG）来聚合客户端模型更新。传统 AFL 采用文献[23]的加权平均聚合规则（公式 19），其权重固定为客户端数据量比例。本文提出可学习聚合机制：通过在服务器端引入代理数据集 P，优化聚合权重λ_i和收缩因子γ（公式 11）。通过梯度下降法联合优化$\{\lambda,\gamma\}$，使聚合后的全局模型在 P 上最小化预测误差。相较于文献[16]的静态权重分配和文献[23]的数据量加权，在 Non-IID 场景下模型准确率有所提高。

在集中聚合阶段，所有客户端的模型参数被发送到中央服务器进行聚合，生成一个全局模型。假设第 i 个客户端的模型参数为w_i^t，并且客户端 i 上的数据量为n_i，全局模型更新的解析解的公式定义为如下：

$$w_g^{t+1} = \frac{\sum_{i=1}^{N} n_i w_i^t}{\sum_{i=1}^{N} n_i} \#(20)$$

这种加权平均方式保证了全局模型能够根据各个客户端的数据量来调整更新权重，从而更精确地反映全局数据分布。然后是关于全局损失函数的定义，其公式如下：

$$F(w) = \sum_{i=1}^{N} \frac{|D_i|}{|D|} L_i(w) \#(21)$$

根据公式（21）的定义，$\frac{|D_i|}{|D|}$是基于客户端数据量的加权系数，使得数据量大的客户端对全局模型的影响更大，这种策略能够提高模型对数据分布不均匀情况的适应性，全局损失函数为所有客户端的数据分布和数据量进行了加权，确保模型对数据量较大的客户端更具适应性。

在 AFL 中，服务器通过解析解的方式进行聚合，以快速且准确地获得更新后的全局模型。服务器在本地执行加权平均来聚合客户端局部模型参数，公式如下：

$$w_g^t = argminF(w) \approx \sum_{i=1}^{N} \frac{|D_i|}{|D|} w_i^t \#(22)$$

公式（22）确保聚合过程能够反映不同客户端的贡献。避免了迭代求解的过程，能够在单步中完成全局模型的更新，从而大幅提升了聚合效率。

服务器使用加权聚合的结果作为新的全局模型参数$w^f(t)$进行更新全局模型，公式如下：

$$w_g^t = \sum_{i=1}^{N} \frac{|D_i|}{|D|} w_i^t \#(23)$$

通过（23）服务器能获得基于各客户端局部训练结果的全局模型更新，更新后的全局模型$w^f(t)$将作为下一轮客户端训练的初始化参数分发至各客户端。并且避免了迭代优化的过程，实现快速聚合。

尽管AFL相比于传统的联邦学习在效率和准确性上有显著提升，但在实际应用中，仍然面临一些关键技术挑战。AFL同样面临客户端分布不均带来的挑战，数据的异质性可能会导致全局模型聚合困难、模型收敛不稳定。为了解决此问题，本文引入自适应加权聚合，在解析性聚合过程中，根据客户端的模型更新质量或数据分布的差异，自适应调整客户端权重，以减少数据异质性对全局模型更新的影响。因此，本文提出的结合可学习聚合权重和解析训练的LAW-AFL框架，旨在提高模型更新对噪声和数据异质性的鲁棒性。

3.2.3　差异化噪声添加策略

为应对联邦学习中潜在的隐私泄露风险（如通过模型参数反推原始数据），本文引入差异化的噪声添加策略。该策略的核心在于根据客户端数据分布的异质性动态调整噪声强度，从而在保护隐私的同时最小化对模型性能的影响。

定义 2（差异化噪声添加与差分隐私保护策略）：设客户端 i 的本地模型更新为Δw_i^t，其数据分布的异质性由 Dirichlet 参数α表示。噪声强度σ

定义为：

$$\sigma = \sigma_{base} \cdot \frac{\sigma_{ref}}{\alpha + \epsilon} \#(24)$$

其中σ_{base}为基础噪声强度，σ_{ref}为参考异质性阈值，ϵ为平滑因子。数据分布越不均匀，即α越小，噪声强度σ越大，反之则减少噪声添加量。

如果出现攻击者伪造指标，就可能操纵权重，从而影响全局模型。例如：恶意客户端可能提交虚假的高质量数据评估结果，或者生成看似高性能的模型参数，从而在聚合过程中获得更高的权重，进而污染全局模型。

因此，提升 LAW- AFL 的安全性以确保数据安全至关重要。在添加噪声的同时引用差分隐私保护策略。改进的噪声添加机制：

$$\Delta w_i^t = w_i^t + N(0, \sigma^2 I) \#(25)$$

其中，噪声强度σ定义为：

$$\sigma = \sigma_{base} \cdot \frac{\sigma_{ref}}{\alpha + \epsilon} \cdot \frac{\Delta}{\epsilon} \#(26)$$

通过 Dirichlet 参数α来衡量数据标签分布的不均衡性。根据公式（25）计算出客户端对应的噪声强度σ_i，利用该噪声强度，生成服从高斯分布$\mathrm{N}(0, \sigma^2)$的噪声ϵ_i。客户端在完成本地模型训练后，获得模型更新量Δw_i^t，将生成的噪声ϵ_i与模型更新叠加，得到噪声扰动后的更新：$\Delta w_i^t + \epsilon_i$。这一操作确保上传到服务器的更新中不仅包含有效的梯度信息，同时掩盖了原始数据中可能泄露的隐私信息。

在数据分布不均匀时，增加噪声以提高隐私保护；在异质性较低时减少噪声，能够在一定程度上保持模型性能，同时兼顾差分隐私保护。

3.3 绝对聚合规则

在聚合阶段，本文引入绝对聚合定律（Absolute Aggregation Law，AA），用于优化大规模客户端场景中的模型参数聚合。该方法基于矩阵伪逆分块法则，通过在全局优化框架下推导出单次聚合策略，从而有效缓解传统方法在异质性数据分布下的性能瓶颈。本文的推导受到了 MP[32]的启发。

定义 3：假设矩阵 X 和 Y 按行分块为上下两部分：

$$X = \begin{bmatrix} X_u \\ X_v \end{bmatrix}, Y = \begin{bmatrix} Y_u \\ Y_v \end{bmatrix} \#(27)$$

其中X_u和X_v均为具有满列秩的子矩阵。根据伪逆分区定理，矩阵X^+的伪逆可以表示为：

$$X^\dagger = \left[\tilde{U}\tilde{V}\right] \#(28)$$

分块伪逆计算：

$$\tilde{U} = X_u^\dagger - R_u C_v (C_u + C_v)^{-1} C_v X_u^\dagger,$$
$$\tilde{V} = X_v^\dagger - R_v C_u (C_u + C_v)^{-1} C_u X_v^\dagger \#(29)$$

并且

$$C_u = X_u^T X_u, C_v = X_v^T X_v,$$
$$R_u = C_u^{-1}, R_v = C_v^{-1} \#(30)$$

全局权重聚合规则：设矩阵 X 和 Y 按行分块为上下两部分：$X = \begin{bmatrix} X_u \\ X_v \end{bmatrix}, Y = \begin{bmatrix} Y_u \\ Y_v \end{bmatrix}$

其中X_u和X_v具有完整的列等级。定义分块伪逆：

$$\widehat{W}_u = X_u^\dagger Y_u, \ \widehat{W}_v = X_v^\dagger Y_v$$

则联合伪逆$\widehat{W} = X^\dagger Y$可表示为：

$$W = \widetilde{W}_u W_u + \widetilde{\mathrm{W}}_v W_v \#(31)$$

其中$\widehat{W}_u$和$\widehat{W}_v$由分块矩阵的协方差矩阵$C_u = X_u^\top X_u$和$C_v = X_v^T X_v$计算得到。

由于

$$\begin{cases} A_u = I - R_u C_v - R_u C_v (C_u + C_v)^{-1} C_v \\ A_v = I - R_v C_u - R_v C_u (C_u + C_v)^{-1} C_u \end{cases} \#(32)$$

因此

$$\begin{cases} C_u = X_u^T X_u \\ C_v = X_v^T X_v \end{cases} \begin{cases} R_u = C_u^{-1} \\ R_v = C_v^{-1} \end{cases} \#(33)$$

全局权重聚合规则揭示了联邦学习中多客户端模型参数聚合的本质：全局模型的权重矩阵W可以通过分块伪逆运算，将各客户端本地训练的权重分量（W_u，W_v）按贡献系数（$\widetilde{W}_u$，$\widetilde{\mathrm{W}}_v$）线性组合而成。直接通过矩阵运算实现全局模型更新，避免传统方法的多轮迭代，减少通信开销；贡献系数$\widetilde{W}_u$，$\widetilde{\mathrm{W}}_v$动态反映客户端数据分布差异，抑制低质量更新的负面影响。

尽管 AA 规则承认两个客户端之间的绝对聚合（即$\widehat{W}_u$和$\widehat{W}_v$），但这种模式可以轻松地广播到多客户端场景。为了详细说明，将$\widehat{W}_{agg,k-1}$表示为聚合了 $k-1$ 个客户端的累积聚合(Accumulated Aggregation，AcAg)全局权重矩阵。通过重写公式(19)（20），下一个带有$\widehat{W}_k (i = 1,2,\ldots,K)$的聚合读取为

$$\widehat{W}_{agg,k} = A_{agg} \widehat{W}_{agg,k-1} + A_k \widehat{W}_k \#(34)$$

令 $C_u \to C_{agg,k-1}, C_v \to C_k$ 。则有 $C_{agg,k} = C_{agg,k-1} + C_k$，其中$C_{agg,k-1}, C_k$是累积协方差矩阵，

因此

$$\begin{cases} A_{agg} = I - C_{agg,k-1}^{-1} C_k \left(C_{agg,k-1} + C_k\right)^{-1} \\ A_k = I - C_k^{-1} C_{agg,k-1} \left(C_{agg,k-1} + C_k\right)^{-1} \end{cases} \#(35)$$

$$\begin{cases} C_{agg,k} = C_{agg,k-1} + C_k = \sum_{i=1}^{k} C_i \\ C_i = X_i^T X_i \end{cases} \#(36)$$

联合训练的权重$\widehat{W} = \widehat{W}_{agg,k}$是以成对的方式聚合各个客户端而产生的。有趣的是，最佳聚合实际上是分别由W_{agg}和W_k加权的两个矩阵（例如$\widehat{W}_{agg,k-1}$和$\widehat{W}_k$）之间的线性组合。聚合不一定遵循从 1 到 k 的顺序索引。可以随机采样可用客户端以与 AcAg 权重进行聚合。权重矩阵中的元素在某种程度上是可以互换的。

3.4 自动加权聚合的解析性联邦学习

在标准AFL的全局聚合步骤中，本文采用了加权平均的方式聚合各个客户端的模型参数。然而，在实际应用中，客户端的数据量和数据质量可能会有所不同，自动加权聚合算法通过自适应调整聚合权重μ_i，可以更加灵活地处理客户端差异，提高聚合的鲁棒性。在本文提出的LAW-AFL算法中，聚合权重μ_i被分解为收缩因子γ，和相对权重λ_i，从而实现更加细致的权重控制。正如算法3所描述的，在指定通信间隔τ时，客户端将局部更新的模型参数w_i^t和权重μ_i发送到服务器。服务器使用μ_i进行加权平均聚合，以解析解的方式更新全局模型w_g^t，避免多次迭代，提高聚合效率。

自动加权聚合的解析性联邦学习算法见算法3

算法 3. 自动加权聚合的解析性联邦学习

输入:客户端集合N，通信伦次T，客户端本地训练周期E，服务器端权重学习周期E_s,初始全局模型w_g^1;每个客户端的数据集D_k;聚合权重参数γ

输出:最终全局模型w_g^t

1：初始化：全局模型w_g^1；聚合统计量矩阵$W_{agg,0} = 0$和$C_{agg,0} = 0$

2：设置聚合权重参数γ, 设置学习率η

3：FOR t=1,...,T DO

//客户端更新阶段

4：分发全局模型:服务器将全局模型w_g^t广播到 n 个客户端.

5:本地模型初始化:每个客户端 i 初始化本地模型;$w_i^t \leftarrow w_g^t$

6：本地训练:每个客户端使用本地数据$D_i = \{X_i, Y_i\}$和 E 轮梯度下降更新本地模型;$w_i^t \leftarrow w_i^t - \eta \nabla L_i(w_i^t)$

7：其中, $L_i(w_i^t)$是客户端 i 的本地损失函数

//计算本地统计信息

8：客户端基于其本地数据集计算以下矩阵:$\widehat{W}_k^T = X_k^T Y_k$, $C_k^T = X_k^T Y_k + \gamma I$

9：在客户端上传参数前添加噪声：

$$\sigma = \sigma_{base} \cdot \frac{\sigma_{ref}}{\alpha + \epsilon} \cdot \frac{\Delta}{\epsilon}$$

10：上传到服务器:客户端将更新后的模型参数w_k^t，统计矩阵$\widehat{W}_k^T$和C_k^T发送到服务器

//服务器更新阶段

11：接收客户端数据:服务器从 m 个选定的客户端接收本地模型$\{w_i^t\}_{i=1}^m$和统计信息$\{\widehat{W}_k^T, w_k^t\}_{i=1}^m$.

12：初始聚合权重λ_i和收缩因子γ

$$\lambda_i = \frac{|D_i|}{|\mathrm{P}|}$$

13：优化聚合权重：在代理数据集 P 上运行 E 轮优化，调整聚合权重

$$\{\lambda, \gamma\} \leftarrow \{\lambda, \gamma\} - \eta \nabla L_{proxy}\left(\{\lambda, \gamma\}\right)$$

14：更新全局模型：使用优化后的权重λ_i，将客户端模型聚合为新的全局模型

$$w_g^{t+1} = \sum_{i=1}^{m} \lambda_i w_i^t$$

//更新全局统计信息

15：累积客户端的统计信息：

$$\widehat{W}_{agg,k} = \widehat{W}_{agg,k-1} + \widehat{W}_k^T$$

$$C_{agg,k} = C_{agg,k-1} + C_k^T$$

16：基于累积的统计信息恢复最终的全局模型权重矩阵：

$$W = \widehat{W}_{agg,k}, C = C_{agg,k}$$

17：更新全局模型：$w_g^{t+1} = WC^{-1}$

18：当通信伦次 T 完成时，服务器输出最终的全局模型：w_g^t

19：END FOR

在客户端本地阶段，每个客户端除了进行传统的本地模型更新外,还需额外计算统计信息（如权重矩阵$\widehat{W}_k$和协方差矩阵C_k），以捕捉本地数据的分布特征，这些统计信息将用于增强全局模型的聚合效果。在服务器端更新阶段，聚合模型的过程分为两步：（1）基于公共数据集优化聚合权重，通过动态调整权重使其适应客户端数据的异构性，提高聚合模型的鲁棒性；（2）在客户端上传参数前添加噪声,结合客户端上传的统计信息，通过矩阵计算方式进一步增强全局模型对全局数据分布的建模能力。总体而言，AFL 提供的矩阵

统计信息有助于建模全局数据分布特征，而FedLAW的动态聚合权重优化则有效应对了客户端间数据分布的异质性问题，两者结合能够显著提升联邦学习的性能与适应性。在客户端训练阶段，各客户端可以使用DNN、ResNet或MLP等架构作为特征提取器。结合自动加权聚合策略，通过收缩因子γ和相对权重λ的动态学习，实现动态调整客户端的聚合贡献，使得AFL能够更好地适应数据异质性，并在聚合时更精确地反映各客户端的贡献。通过解析解的方式计算全局模型，避免了迭代优化过程，大幅提升聚合阶段的计算效率。这些网络架构在预训练后能够有效提取图像或复杂数据的高层次特征，从而加速训练过程，并减少客户端计算负担。

4 实验

4.1 实验设置

（1）数据集

为了评估LAW-AFL方法的有效性,本文通过在四个不同的联邦基准数据集上进行了实验,即MNIST、FashionMNIST、CIFAR-10和CIFAR-100上评估了本文方法和基准方法，数据集的统计信息如表2所示。数据集的训练数据按照一定策略分配给每个客户端，每个客户端都有自己的训练集。

表2 数据集统计信息

数据集	训练样本	测试样本	特征	类别
MNIST	60000	10000	$28\times28\times1$	10
Fashion-MNIST	60000	10000	$28\times28\times1$	10
CIFAR-10	50000	10000	$32\times32\times3$	10
CIFAR-100	80000	20000	$32\times32\times3$	100

数据划分按照文献[33]方法对$p_c \sim Dir_i(\alpha)$进行采样，并按照$p_{c,i}$的比例随机分配c类训练数据给客户端i。Non-IID(α)用于表示这种模拟方法，这里的α与Dieichlet分布中的α相同，它作为控制参数，用于调整数据分布的不平衡程度。通过调整Dieichlet参数α来控制数据分布的异质性程度。α值越小，表示客户端之间的数据分布差异越大，即Non-IID程度越高，反之，α值越大，表示数据分布越接近，即IID程度越高。在本文中客户端数量为100，每轮随机选择10个客户端参与训练。在MNIST数据集下每个客户端300-700张图象，CIFAR-10客户端拥有400-800张数据集，CIFAR-100数据集下客户端用于800-1200张图像。

本文采用训练迭代次数（Training Iterations）作为衡量指标，记录在单次全局通信轮次内，客户端本地进行的多步训练迭代，以验证模型在单周期内的收敛性。这种更为精细的单周期优化方法，有助于进一步提升模型性能。

为了评估LAW-AFL性能的稳定性，在相同实验条件下独立重复运行010次，记录通信轮次上的准确率并计算标准差σ，阴影区域表示平均值$\pm1\sigma$，用于直观反映模型性能在重复实验中的波动程度。

（2）实验评价指标以及基准设置

本文采用最常用的精确度评价指标：准确率（accuracy，ACC）和Top1 Hessian特征值以及训练时间（Training Time，T）。准确率是衡量分类任务中正确预测样本比例的关键指标，对比不同模型在相同数据集上的准确率。更高的准确率通常表明模型具有更好的性能。Top-1 Hessian特征值是衡量模型优化稳定性的重要指标，反映模型优化曲面的陡峭程度，值越大表明优化过程越不稳定。对于损失函数L_{proxy}，其Hessian矩阵计算为：

$$H = \frac{\vartheta^2 L_{proxy}}{\vartheta\theta\vartheta\theta^T} \#(37)$$

在全局模型w_g^T处采样小批量数据，然后利用自动微分框架计算二阶导数；最后迭代求解Hessian矩阵得最大特征值。

较低的Top1 Hessian特征值通常表明模型优化过程更平稳，梯度更新更稳定，从而有助于提高模型的鲁棒性，其增长速度和最大值反映了模型优化过程是否稳定，用以分析模型在训练过程中是否过度拟合局部数据。此外，训练时间作为衡量算法效率的重要指标，在联邦学习中，通信成本和训练效率是重要的关注点。更短的训练时间表示模型具有更高的计算效率和更低的通信开销，特别是在大规模客户端环境中具有重要意义。

在基准设置方面，算法LAW-AFL的主干模型源自FedLAW，这一选择基于其在联邦学习领域的卓越表现。为了更全面地验证LAW-AFL的性能，本文参考了文献[16]中的实验对比结果，该文献已经对FedLAW与多种基准方法进行了详尽的对比，包括FedDF[6]、FedAVG[2]、FedPROX[14]、

FedDYN[15]以及 FedBE[34]等。这些基准方法各自具有独特的特点和优势，为评估 LAW-AFL 提供了有力的参照。

在此基础上，进一步设立了 FedLAW 和 AFL 两个基线模型，以便与 LAW-AFL 进行更为深入的对比。FedLAW 作为主干模型，其性能表现将为本文提供重要的参考基准。而 AFL 则是一种新兴的联邦学习框架，通过引入自动学习的机制来提高模型的性能[35]。将这两个基线模型纳入对比范围，有助于更全面地评估 LAW-AFL 在结合 FedLAW 和 AFL 优势方面的表现，以及其在处理联邦学习场景中的复杂性和挑战时的能力。

（3）模型架构及设置

在深度学习领域，采用了六种具有代表性的模型架构——CNN、DNN、ResNet20、ResNet50、DenseNet121 以及多层感知机（MLP），选择不同架构的原因在于它们在结构设计、特征提取和学习能力上存在显著差异，这有助于全面评估和验证所提出方法的适用性与鲁棒性。在实验设置中，设计了多种模型架构与参数设置。对于 MLP，输入为图像展平后的向量（如 MNIST 为 784 维），包含两个全连接隐藏层，神经元数分别为 512 和 256，采用 ReLU 激活函数，输出层为 Softmax 分类层。基础卷积网络（CNN）的结构包括：第一卷积层使用 32 个 5×5 卷积核；第二卷积层使用 64 个 5×5 卷积核；随后是全连接层（1024 神经元）和 Softmax 输出层。ResNet 架构中，ResNet-20 由 3 个阶段组成，每个阶段包含 3 个残差块（共 20 层），初始卷积层为 3×3 卷积（64 通道）；ResNet-50 采用 Bottleneck 残差块构成 4 个阶段（共 50 层），初始卷积层为 7×7 卷积（64 通道），DenseNet-121 由 4 个密集块组成，每个过渡层包含卷积和池化操作。

在不同的模型架构下，分别进行了分析和对比实验，通过对比不同模型在相同任务下的准确率、训练时间等关键性能指标，可以直观地评估各模型的性能优劣。

4.2 实验结果

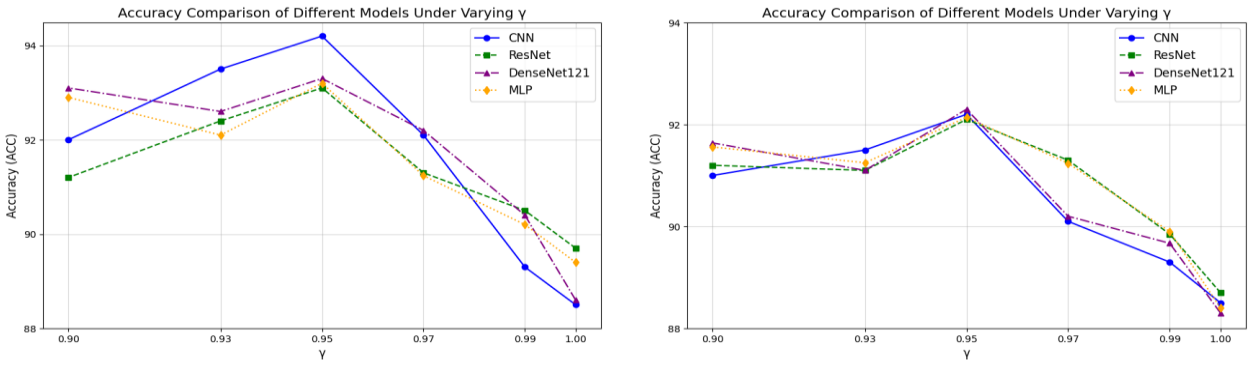
为了验证 LAW-AFL 的性能，设计实验将其与基线模型进行了对比。

4.2.1 不同γ对模型性能影响

在数据集上，通过调整收缩因子γ来最小化全局模型的损失函数，因此在各模型上通过调整γ观察其准确率，找到最优的收缩因子γ。

在 MNIST 和 CIFAR-10 两个不同数据集上，不同的γ值对模型的表现产生了不同的影响。通过图 1 可知，γ值较小（γ=0.95）时，模型泛化能力显著提升，这一发现与以往的理论有所不同。在$\gamma = 0.95$时，模型的准确率达到了最佳水平。由此可见，在给定的实验设置下，存在一个最优的γ值，它能够平衡正则化和优化的需求，偏离该值可能导致性能下降。

为了进一步评估不同γ设置下对模型的影响，我们采用四种不同结构的模型进行了对比实验。如表 3-表 6 所示，γ=0.95 是在标准数据集和模型架构下的经验最优值。Non-IID 程度越高，需更小的γ以增强正则化，复杂的模型对γ更敏感，需动态调整；分类任务中γ通常稳定在 0.93-0.97，此外基于梯度下降优化方法其最优γ区间分布在 0.94-0.95 之间。在准确率损失不超过 0.5%的条件下，手动调参可节省 22.4%的训练时间，更适合资源受限场景。因此，后续对比实验均在这一最优设置下进行。为了验证方法在资源受限环境下的可行性，本文在两种 Non-IID 划分的客户端场景中，以相同的训练预算对各模型性能进行了对比测试。这两种设置会对联邦学习的模型训练过程产生不同的影响，在这种情况下，证明模型处理客户端之间异质性的性能。



（a）MNIST （b）CIFAR-10

图 1 不同γ值下的模型架构准确率对比

表 3 在 MNIST 数据集下，对于不同γ设置下对各模型效用性的影响

模型	γ =0.90	γ = 0.93	γ = 0.95	γ = 0.97	γ = 0.99	γ = 1.00
CNN	92.34	93.87	**94.95**	92.12	89.54	88.94
ResNet-20	91.72	92.67	**93.76**	91.64	91.36	89.86
DenseNet121	93.56	92.71	**93.83**	92.13	91.25	88.93
MLP	92.64	92.65	**93.81**	91.51	90.99	89.34

表 4 在 FashionMNIST 数据集下，对于不同γ设置下对各模型效用性的影响

模型	γ =0.90	γ = 0.93	γ = 0.95	γ = 0.97	γ = 0.99	γ = 1.00
CNN	92.46	**93.57**	93.53	93.15	93.06	92.97
ResNet-20	91.57	92.19	92.39	**92.46**	92.16	92.04
DenseNet121	92.56	92.84	**93.49**	93.12	92.98	93.05
MLP	91.67	91.49	**92.67**	92.13	92.08	92.57

表 5 在 CIFAR-10 数据集下，对于不同γ设置下对各模型效用性的影响

模型	γ =0.90	γ = 0.93	γ = 0.95	γ = 0.97	γ = 0.99	γ = 1.00
CNN	91.58	91.36	**92.28**	91.58	89.57	88.64
ResNet-20	90.69	91.04	**92.14**	90.67	89.94	88.76
DenseNet121	90.57	91.46	**92.29**	90.24	89.59	88.51
MLP	91.56	91.04	**92.25**	90.06	89.35	88.32

表 6 在 CIFAR-100 数据集下，对于不同γ设置下对各模型效用性的影响

模型	γ =0.90	γ = 0.93	γ = 0.95	γ = 0.97	γ = 0.99	γ = 1.00
CNN	93.16	92.43	**93.58**	93.12	89.54	89.46
ResNet-20	92.49	93.19	**93.61**	92.97	91.01	89.16
DenseNet121	92.87	**93.49**	93.21	93.04	90.76	90.27
MLP	92.19	88.57	92.69	**93.24**	91.57	90.34

4.2.2 可学习聚合权重的有效性分析

为了评估可学习聚合权重与固定权重对模型准确率的影响，我们在数据集为 CIFAR-10，模型架构为 ResNet-20，$\alpha = 0.1$的设置下将 LAW-AFL 与三种模型进行了对比实验。实验结果如表 7 所示，LAW-AFL（可学习聚合权重）的准确率达到 87.4%，相比于固定权重 LAW-AFL 高 9.1%，与其他基线方法相比，LAW-AFL 动态优化能更好适应客户端数据分布的差异性。

表 7 可学习权重与固定权重下模型准确率对比

方法	准确率（%）	收敛轮次	客户端权重方差（$\times 10^{-2}$）
LAW-AFL（可学习权重）	**87.4**	**120**	**3.2**
LAW-AFL（固定权重）	78.3（-9.1）	190	0.0
FedLAW（静态优化权重）	82.1（-5.3）	160	1.8
FedAVG	76.5	200	-
FedPROX	77.4	195	-

在全局通信轮次方面，LAW-AFL 仅需 120 轮便实现模型收敛，而 LAW-AFL（固定权重）和 FedLAW 静态优化权重分别需要 190 轮和 160 轮，其他基线方法普遍需要超过 165 轮才能收敛。这表明动态聚合权重能够有效减少低质量客户端的干扰，加快整体收敛过程。从客户端权重方差来看，LAW-AFL 的方差为 3.2，更高的方差表明 LAW-AFL 对不同数据质量及模型性能作出差异化处理。

LAW-AFL 通过动态学习聚合权重，有效压制了恶意客户端的影响，实验结果如表 8 所示，LAW-AFL（可学习权重）其平均权重仅为 1.3，而固定权重的 LAW-AFL 恶意客户端平均权重为 4.3。FedLAW 通过静态优化聚合权重，表现比固定权重下的 LAW-AFL 好，其恶意客户端平均权重为 2.6。而 FedAVG 的固定权重策略下恶意客户端平均权重高达 8.6。FedPROX 虽通过正则项部分抑制了异常客户端，但其权重仍较高。LAW-AFL 在 Non-IID 环境下保持了 90.1%的高准确率，通过动态优化客户端权重，LAW-AFL 不仅能够针对性地削弱恶意客户端的影响，还能更好地适应客户端数据的差异性。

表 8 MNIST，$\alpha = 0.1$设置下权重有效性

方法	准确率（%）	恶意客户端平均权重（$\times 10^{-2}$）
LAW-AFL（可学习权重）	**90.1**	**1.3**
LAW-AFL（固定权重）	84.6	4.3（-3.0）
FedLAW	86.7	2.6
FedAVG	72.5	8.6
FedPROX	78.4	5.2

在 MNIST、$\alpha = 0.1$设置下，我们验证了可学习聚合权重对 Hessian 特征值的影响。实验结果由表 9 可知，LAW-AFL 通过动态学习聚合权重，使模型的 Top-1 Hessian 特征值显著降低至 1.2，相比固定权重策略和其他基线方法均有大幅改善。较低的 Hessian 特征值表明模型收敛曲面更平缓，优化过程更为稳定，有助于避免梯度震荡和局部发散。

LAW-AFL（可学习聚合权重）的值仅为 0.8，相较于 LAW-AFL 固定权重（2.5）和其他基线方法（2.0–2.6）大幅降低。

表 9 可学习聚合权重对 Hessian 特征值的影响

方法	Top-1 Hessian 特征值（$\times 10^{-3}$）	梯度方差（$\times 10^{-3}$）
LAW-AFL（可学习权重）	**1.2**	**0.8**
LAW-AFL（固定权重）	3.7	2.5
FedLAW	2.1	1.4
FedAVG	3.8	2.6
FedPROX	3.6	2.4

因 LAW-AFL 及 FedAVG 等方法采用固定权重，导致 Hessian 特征值和梯度方差偏高，影响了模型优化的平滑性和整体性能。FedLAW 虽然通过公共数据集进行权重优化，但未引入自适应 **Error! Reference source not found.**调整机制，其性能略低于 LAW-AFL。这表明动态聚合权重能够有效抑制不同客户端之间更新方向的冲突。

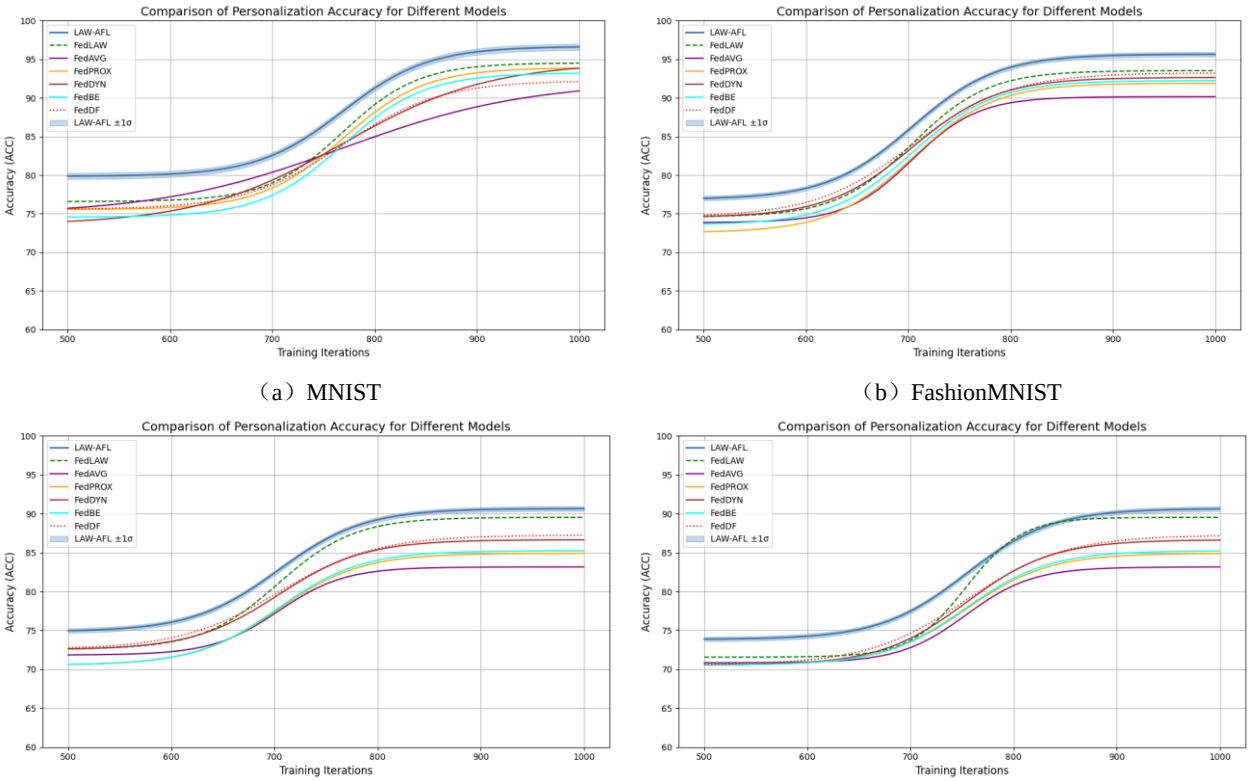
4.2.3 各模型在不同数据集和不同 Dieichlet 参数 α配置下的表现

在 FL 过程中，E 表示每个客户端在与服务器通信前，在本地数据上训练的迭代次数，随后将模型更新发送至服务器进行聚合，生成新的全局模型。

当 E 较大时，客户端在本地训练轮次增加，可能加剧因数据差异带来的模型偏移。α控制数据异质性，E 影响训练深度与同步性，两者共同决定了联邦学习的训练动态。

首先研究在 Non-IID($\alpha = 0.1$)数据分布下，

不同数据集在 DNN 模型架构上的学习曲线的表现，由于 Non-IID 数据分布的挑战和客户端之间的数据分布不均匀，从图 2 中可以看出，在 MNIST 数据集下，LAW-AFL 模型在最终的准确率达到了 90%左右，FedLAW 模型的准确率略高于其他基线模型，准确率达到了 86%左右，阴影区域的宽度反映了模型在不同运行中的波动情况，LAW-AFL 的误差带较窄，多次实验结果波动较小。在 Fashion-MNIST 数据集下，LAW-AFL 呈缓慢且稳定的上升趋势，准确率最后维持在 90%左右，其次是 FedLAW 模型准确率在 86%左右，其他基线模型的准确率集中分布在 83%左右。此外在较为复杂的 CIFAR-10 和 CIFAR-100 数据集上，LAW-AFL 保持较好的性能，其准确率均保持在 89%左右，与其他基线模型相比，LAW-AFL 通过动态调整聚合权重，有效缓解了 Non-IID 环境下客户端数据分布不均的影响。



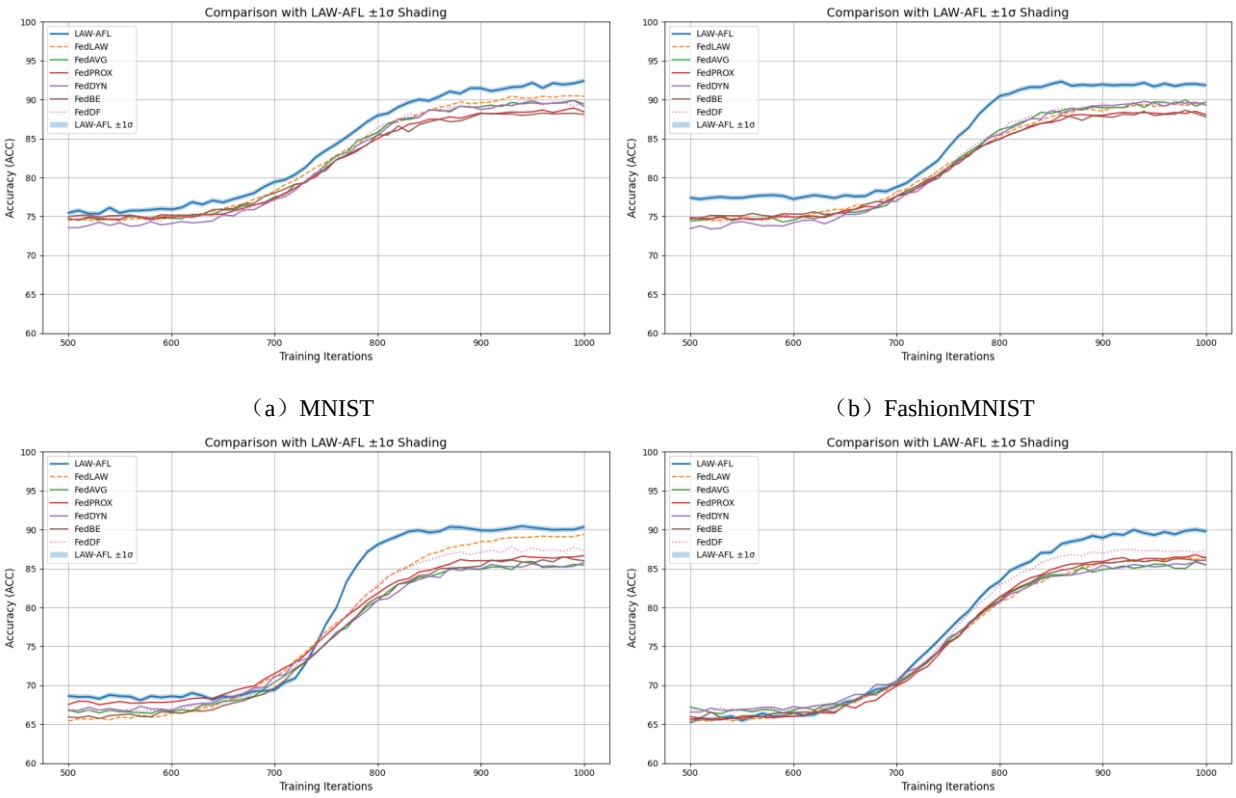
（a）MNIST （b）FashionMNIST

（c）CIFAR-10 （d）CIFAR-100

图 2 在 DNN，Non-IID($\alpha = 0.1$)分布下的学习曲线

在四个数据集上选取 DNN 和 Non-IID($\alpha = 0.05$)作为评估指标，对各模型在不同异质性环境下的适应能力进行分析与验证，实验结果如图 3 所示，基线模型在数据集上的结果略弱于 Non-IID($\alpha = 0.1$)上的结果，但阴影区域相比较图 2 更窄了，模型在经历过多次实验后，表现更加稳定。图 3 设置纵轴为准确率，横轴为客户端本地训练迭代次数（单周期内）。在 Non-IID($\alpha = 0.05$)上的准确率变化很小，并且 LAW-AFL 模型并未出现很大的波动。

LAW-AFL 通过代理数据集驱动的权重分配，在高异质性场景下抑制低质量客户端的负面影响。即使在数据分布极度不均的情况下，模型仍展现出稳定的性能和良好的鲁棒性。以 FedDYN 为例，在 CIFAR-10，Non-IID($\alpha = 0.1$)下，其准确率在 83%左右，而在 Non-IID($\alpha = 0.05$)下，FedDYN 的准确率下降至 81%左右，较小的α值会导致现有的 FL 对应方法的性能下降，而 LAW-AFL 则表现出了较为稳定的性能。



（a）MNIST （b）FashionMNIST

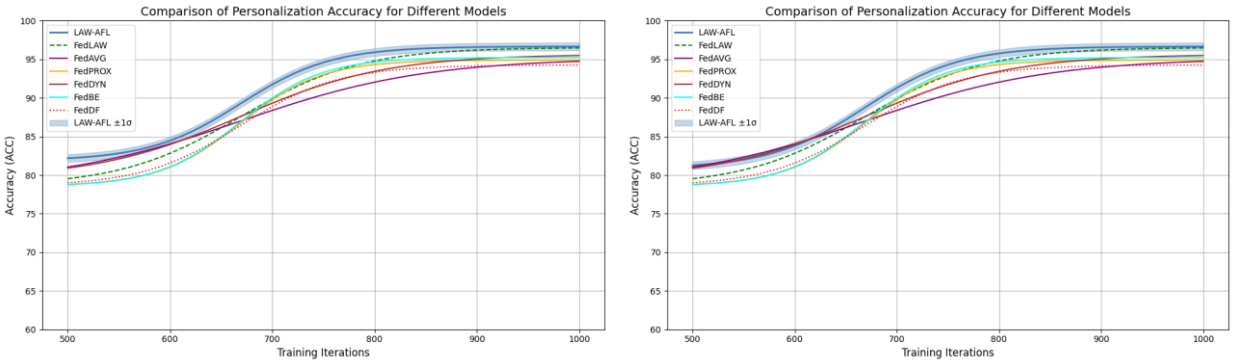
（c）CIFAR-10 （d）CIFAR-100

图 3 在 DNN，Non-IID($\alpha = 0.05$)分布下的学习曲线

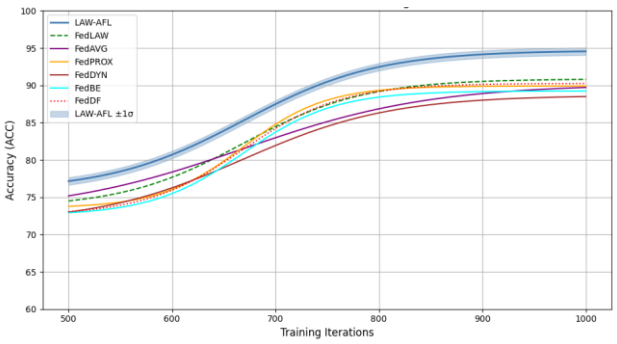
在 ResNet20 结构和 Non-IID($\alpha = 0.1$)数据分布条件下，我们对 LAW-AFL 等七种模型在四个不同数据集上的分类性能进行了评估，分析了其随训练轮次增加的准确率变化趋势，实验结果如图 4 所示。

在 MNIST 数据集上，LAW-AFL 的表现始终优于其他模型，随着训练轮次从 500 增加到 1000，模型准确率稳步上升，最终超过 95%。FedLAW 表现次之，但略低于 LAW-AFL，而其余模型的曲线较为接近，准确率在 80%到 90%之间，波动较小。在较复杂的 FashionMNIST 数据集上，LAW-AFL 通过动态权重优化和收缩因子的稳定性控制取得了较优的性能，准确率最终接近 95%。FedLAW 紧随其后，表现出较高的稳定性，而其他模型的准确率差距较小。在 CIFAR-10 数据集上，LAW-AFL 的学习曲线仍然最为陡峭，准确率在 1000 轮次时突破 90%，显著高于其他模型。FedLAW 的准确率接近 87%，而其余模型（如 FedAVG、FedPROX 等）的表现相对较差，准确率在 85%左右，波动幅度略有增加。

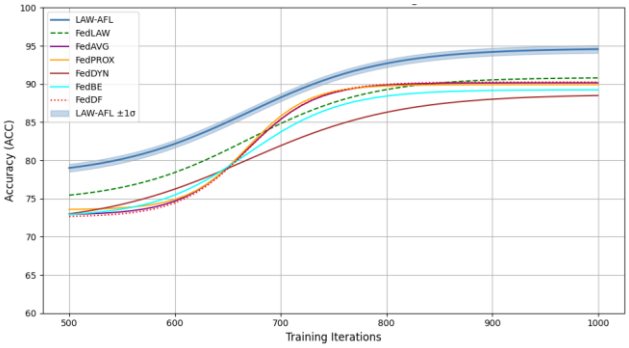
在 CIFAR-100 数据集上，LAW-AFL 最终准确率超过 90%，表现出较强的泛化能力。其余模型的准确率均低于 90%，且在训练后期出现一定程度的波动。



（a）MNIST （b）FashionMNIST



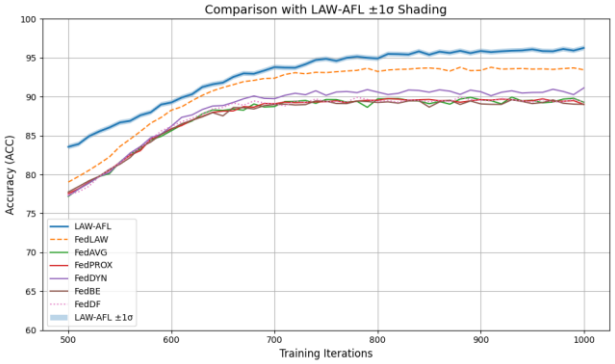
（c）CIFAR-10



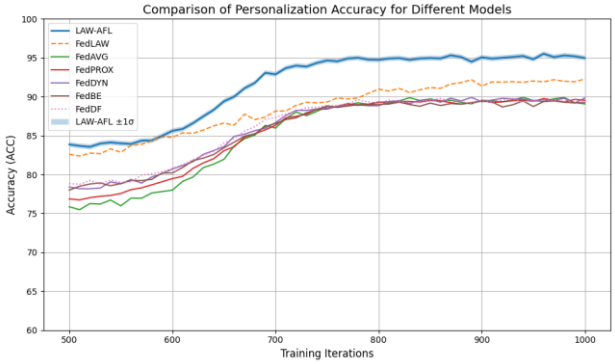
（d）CIFAR-100

图 4　在 ResNet20，Non-IID($\alpha = 0.1$)分布下的学习曲线

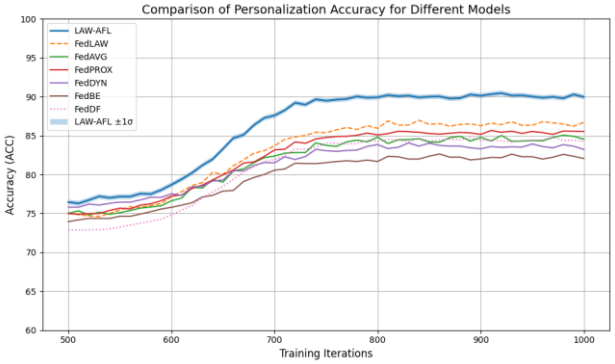
在 ResNet20 架构和 Dirichlet 分布参数为($\alpha = 0.1$)分布下，在$\alpha = 0.1$场景下，数据分布的异质性相对较低，为验证模型训练的稳定性和收敛速度，我们对各模型的表现进行了对比分析，实验结果如图 5 所示，LAW-AFL 的收敛速度和最终准确率略高于$\alpha = 0.05$场景。在$\alpha = 0.05$高异质性分布下，尽管数据分布差异显著，在迭代过程中出现一定的波动，但由于 LAW-AFL 通过收缩因子γ的自适应正则化作用使得在极端 Non-IID 下的 LAW-AFL 波动范围相对比较平稳并且仍能快速收敛，在最终准确率上保持领先。此外，与 FedAVG 和 FedRPOX 其他模型相比，LAW-AFL 的学习曲线在早期阶段呈现更快的上升趋势，这表明 LAW-AFL 的自适应聚合策略在异构数据环境中显著提升了全局模型的训练效果，同时保持了模型性能的稳定性和泛化性。



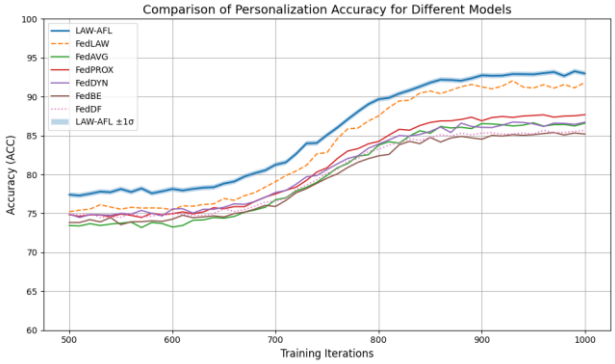
（a）MNIST



（b）FashionMNIST



（c）CIFAR-10

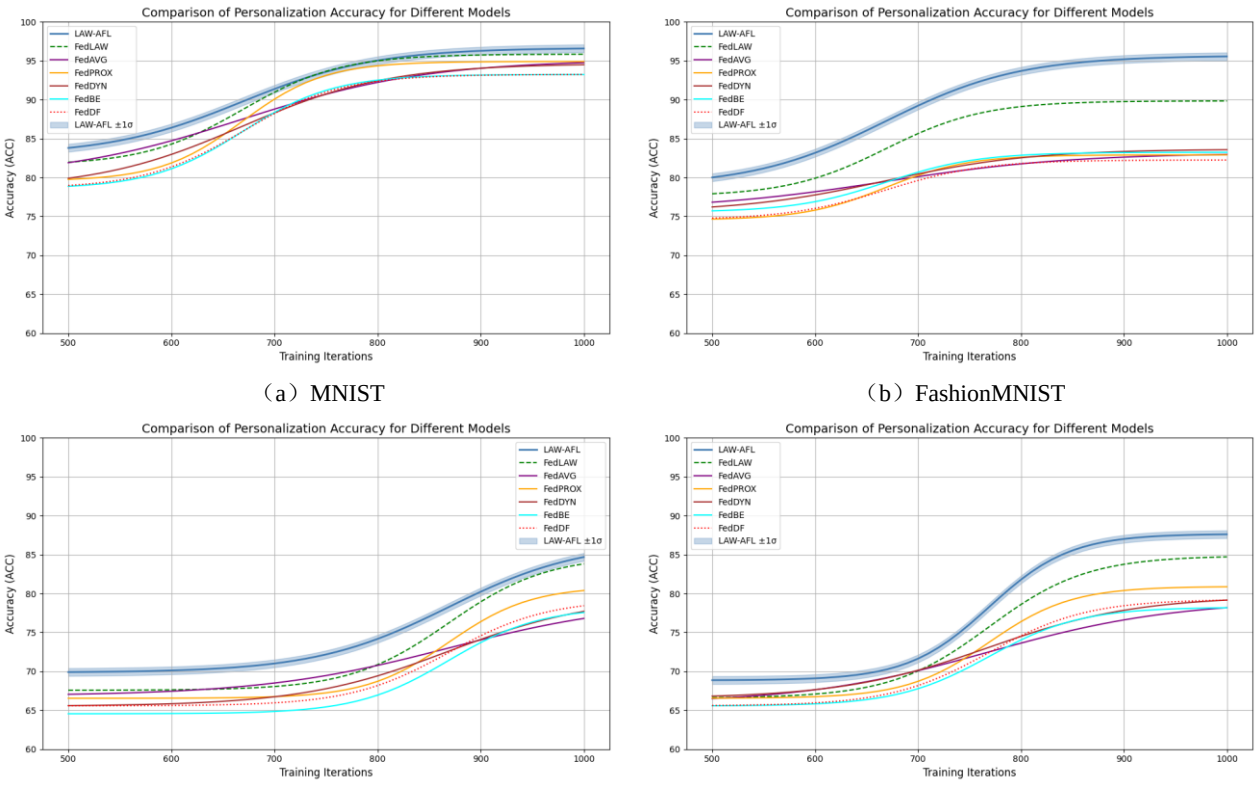


（d）CIFAR-100

图 5　在 ResNet20，Non-IID($\alpha = 0.05$)分布下的学习曲线

在 DenseNet121 模型和 Non-IID($\alpha = 0.1$)数据分布下对多种模型进行了性能评估，实验结果如图 6 所示，各模型在 MNIST 数据集上准确率整体较高，如图 6（a）各方法均能达到 95%左右，

但 LAW-AFL 相较其他方法在后期（900~1000 次迭代）仍有明显优势。由图 6（b）可得出，FashionMNIST 数据集上的准确率显著低于 MNIST 数据集，方法间差异更加显著，LAW-AFL 和 FedLAW 明显优于其他方法，表明在较复杂的视觉任务上，先进方法对性能的提升更加重要。图 6（c）曲线起点较低大约在 65%，表明任务难度提升。LAW-AFL 的准确率提升速度最快。图 6（d）为任务最具挑战性的 CIFAR-100 数据集，所有方法的初始准确率较低大约在 65%，但随着训练进行，LAW-AFL 的曲线更平滑。为了量化模型在相同设置下的稳定性，在数据集上重复 10 次独立实验，在 CIFAR-100 上，阴影带整体比 MNIST 更宽，面对更难的任务时，模型会随之产生较大的波动。



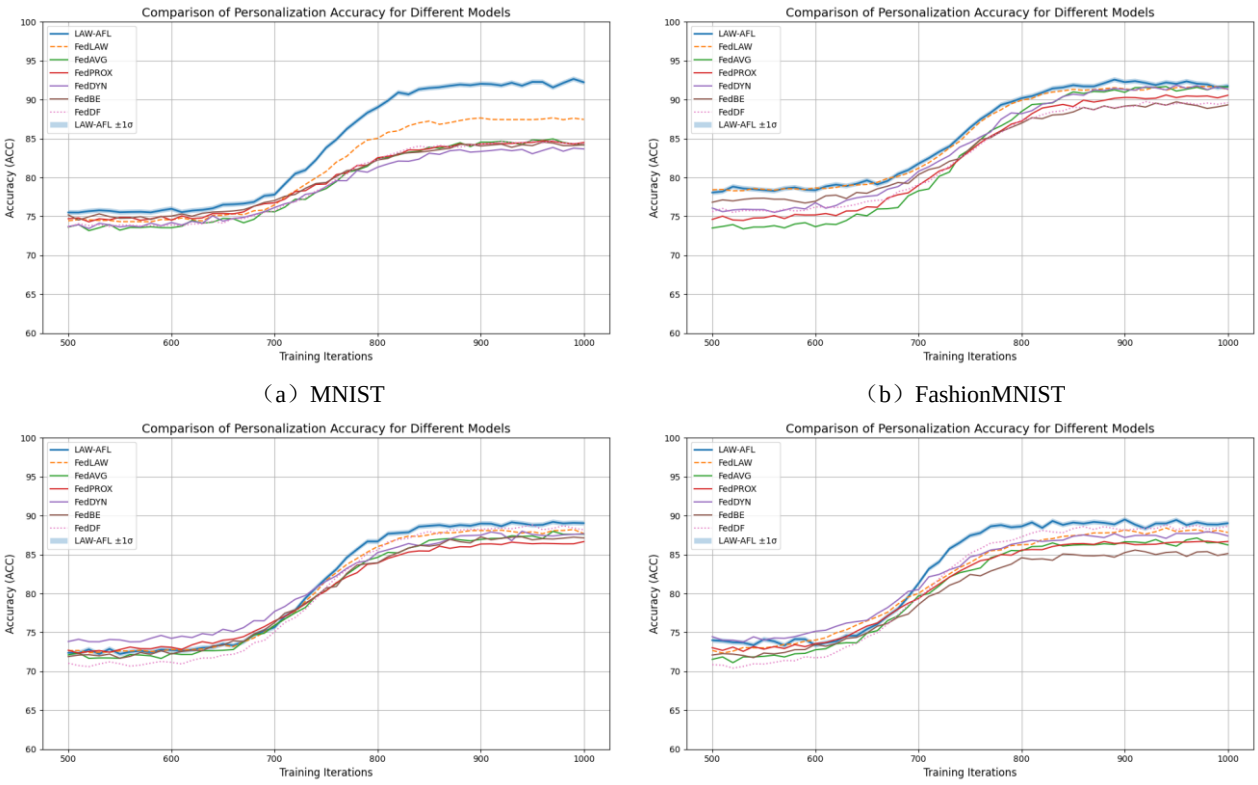
（a）MNIST （b）FashionMNIST

（c）CIFAR-10 （d）CIFAR-100

图 6 在 DenseNet121，Non-IID($\alpha = 0.1$)分布下的学习曲线

在采用 DenseNet121 架构并将 Dirichlet 分布参数设置为 0.05 的条件下，我们在四个数据集上对各模型的准确率进行了评估，实验结果如图 7 所示。LAW-AFL 在 MNIST，$\alpha = 0.1$的非独立同分布数据上准确率达到 92.75%，在 FashionMNIST 上准确率达到了 93.54%。

在 CIFAR-10 与 CIFAR-100 数据集上也达到了 87%左右的准确率，对比$\alpha = 0.1$与$\alpha = 0.05$两组实验，LAW-AFL 模型最终准确率并未出现较大的差异，LAW-AFL 的闭式解聚合基于分块矩阵伪逆在极端数据分布下仍能保持权重分配的数学最优性，而传统迭代方法易受局部梯度噪声影响。因此 LAW-AFL 在各数据集中的性能保持领先，表明使用神经网络进行自动加权聚合的解析性联邦学习模型效果更佳。



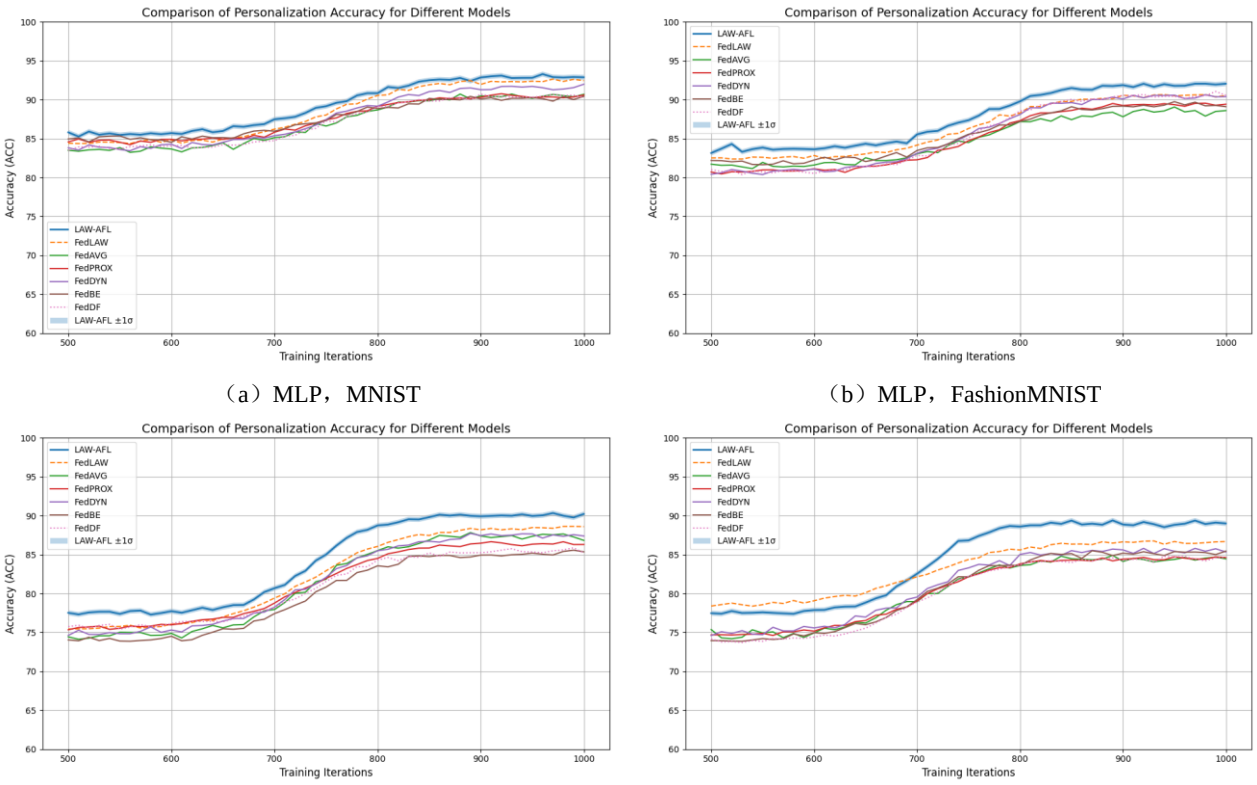
（a）MNIST （b）FashionMNIST

（c）CIFAR-10 （d）CIFAR-100

图 7 在 DenseNet121，Non-IID($\alpha = 0.05$)分布下的学习曲线

在 Non-IID($\alpha = 0.1$)数据分布下选取了 MLP 和 ResNet50 两种模型架构，以评估各联邦学习方法在低异质性环境下的分类性能表现。实验结果如图 8 所示，横轴为训练迭代次数，纵轴为模型的准确率。图 8（a）展示了多层感知机 MLP 在 MNIST 数据集上的性能表现。α越小，客户端本地数据分布越极端且高度一致，不同实验得到的本地分布结构差异反而越小，阴影区域相较于 $\alpha = 0.1$更窄，但准确率降低。

在 MNIST 数据集上，各模型准确率随训练迭代次数的增加均表现出稳步提升。LAW-AFL 显著优于其他方法，准确率最终接近 93%。FedLAW 紧随其后，但略低于 LAW-AFL，表明其自适应聚合策略在数据分布异质性下仍有一定局限性。图 8(b)各模型的准确率提升速度较慢，LAW-AFL 依然保持最优性能，且随着迭代次数增加，准确率持续提升。其他模型的表现相对接近，但与 LAW-AFL 之间存在差距。说明 LAW-AFL 在更复杂的数据分布下仍能适应异质性并提高全局模型的泛化性能。在 CIFAR-10 数据集上，图 8（c）可以看出各模型的准确率提升更加平缓，曲线表现出一定波动。LAW-AFL 显著优于其他方法，并在后期保持稳定增长。FedLAW 和其他方法（如 FedPROX、FedAVG）的准确率较低，表明它们在具有更高维特征的数据集（如 CIFAR-10）上对异质性数据的适应能力不足。图 8（d）展示在 CIFAR-10 数据集上，各模型的准确率提升更加平缓，曲线表现出一定波动。LAW-AFL 在性能上依然占据领先位置，并展现出一定的稳定性。

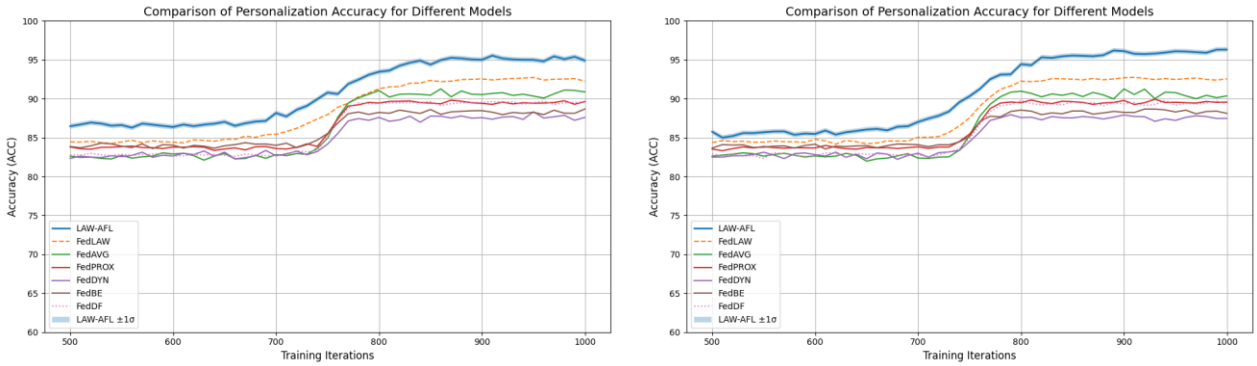


（a）MLP，MNIST　　（b）MLP，FashionMNIST

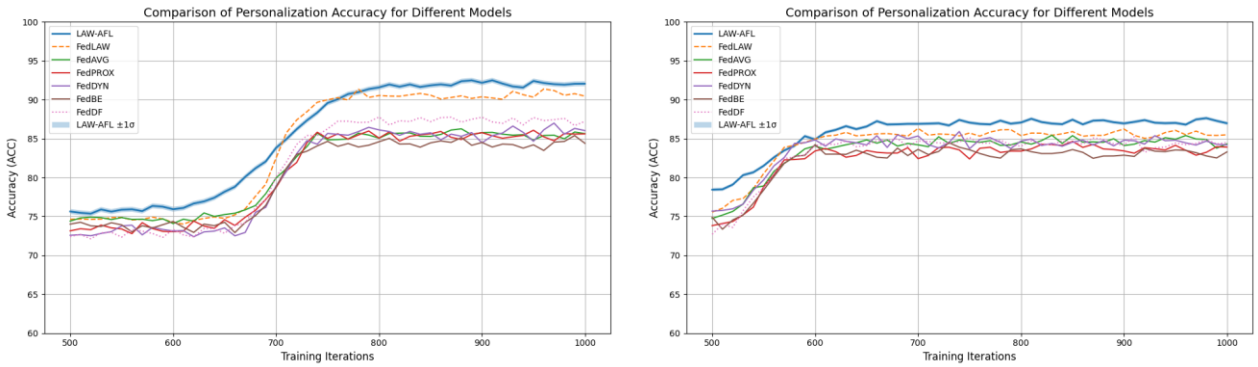
（c）ResNet50，CIFAR-10　　（d）ResNet50，CIFAR-100

图 8　在 Non-IID($\alpha = 0.1$)分布下的学习曲线

在 Non-IID 不同设置下，为了进一步验证 LAW-AFL 在不同数据集和不同模型架构下的性能表现，与 FedDF、FedBE 和 FedAVG 等方法进行了对比实验。实验结果如图 9 所示，LAW-AFL 能更好地利用自适应聚合策略，能够在异构数据环境中显著提升了全局模型的训练效果。LAW-AFL 模型性能超过了基线模型。采用了不同的模型架构，在 MLP 上训练 MNIST 和 FashionMNIST 数据集，LAW-AFL 模型的准确率达到 91.77%左右，在 ResNet50 模型架构下测试了 LAW-AFL，从图中可以看出，模型仍然保持良好的性能，准确率维持在 85%-88%左右，相比较于基线模型，在$\alpha = 0.05$环境下，准确率的波动并未很明显，LAW-AFL 的阴影带更窄，通过多次独立实验仍然保持较稳定的性能。表 10、表 11 为 LAW-AFL 与其他六种方法在不同模型架构上$\alpha = 0.1$和$\alpha = 0.05$时的准确率比较结果。



（a）MLP，MNIST　　（b）MLP，FashionMNIST



（c）ResNet50，CIFAR-10　　（d）ResNet50，CIFAR-100

图 9　在 Non-IID($\alpha = 0.05$)分布下的学习曲线

表 10　$\alpha = 0.1$下具有不同模型架构的四个数据集上比较传统 FL 方法和 LAW-AFL 的 TOP-1 测试准确率

数据集	MNIST		FashionMNIST		CIFAR-10		CIFAR-100	
模型	MLP	LeNet	MLP	LeNet	CNN	ResNet	CNN	ResNet
FedLAW	88.61	88.26	89.29	88.08	80.17	89.46	80.74	88.49
FedAVG	88.11	88.08	87.68	88.61	70.59	88.57	79.52	86.16
FedPROX	89.33	87.01	88.64	87.68	67.66	87.51	78.49	85.34
FedDYN	89.24	89.68	88.47	72.45	66.1	86.65	77.68	85.29
FedBE	89.14	85.96	86.16	85.94	69.60	86.36	78.42	86.28
FedDF	89.09	85.90	86.22	85.10	69.88	86.94	77.73	85.30
LAW-AFL	**90.66**	**90.51**	**90.65**	**88.86**	**82.46**	**90.27**	**82.83**	**89.61**

表 11　($\alpha = 0.05$)时具有不同模型架构的四个数据集上比较传统 FL 方法和 LAW-AFL 的 TOP-1 测试准确率

数据集	MNIST		FashionMNIST		CIFAR-10		CIFAR-100	
模型	MLP	LeNet	MLP	LeNet	CNN	ResNet	CNN	ResNet
FedLAW	85.24	86.11	86.51	85.17	**80.46**	85.62	82.39	84.63
FedAVG	81.46	85.33	80.24	85..48	74.97	83.69	80.45	83.22
FedPROX	80.21	77.68	83.56	87.01	79.21	85.94	82.67	83.76
FedDYN	81.43	78.48	83.49	88.17	77.34	83.06	79.16	85.29
FedBE	79.64	79.12	79.27	77.25	77.61	84.96	80.43	83.47
FedDF	87.31	85.71	87.27	79.84	74.73	85.04	77.24	84.34
LAW-AFL	**89.97**	**89.66**	**90.24**	**88.18**	80.27	**89.23**	**83.49**	**85.19**

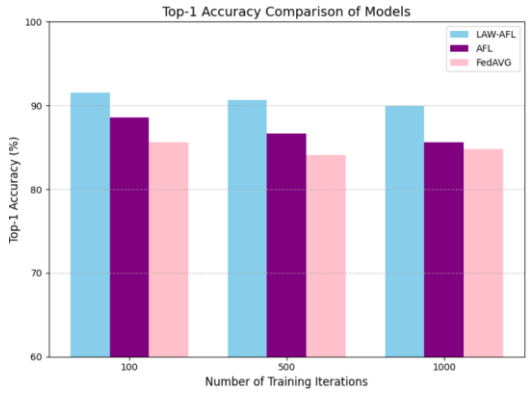
4.2.4　不同客户端参与下的各模型准确率

不同联邦学习模型（LAW-AFL、AFL 和 FedAVG）在不同数据集、数据异质性α和训练迭代次数下的 Top-1 准确率表现。实验结果如图 10 所示，横轴表示训练迭代次数（100、500、1000），纵轴为 Top-1 准确率（%）。由图 10（a）可知，随着训练迭代次数从 100 增加到 1000，所有模型的 Top-1 准确率均有所提升。LAW-AFL 的准确率始终高于其他模型，并在训练迭代次数为 1000 时达到接近 90%，表现出良好的全局优化能力。AFL 的表现次之，但与 LAW-AFL 的差距较明显。FedAVG 的表现最差，准确率徘徊在 85%左右，表明其在低异质性($\alpha = 0.1$)场景下仍然存在局限性。AFL 相较于传统 FedAVG 方法具有一定改进，但在处理异质数据的鲁棒性上仍不及 LAW-AFL。图 10（b）在 CIFAR-100 这一复杂数据集上，所有模型的 Top-1 准确率整体表现较低，且差距较明显。LAW-AFL 的表现依然最优，随着训练迭代次数增加，准确率显著提升，1000 次迭代时准确率约 86%。AFL 的准确率在 100 次迭代时略低于 LAW-AFL，后续提升幅度较小。

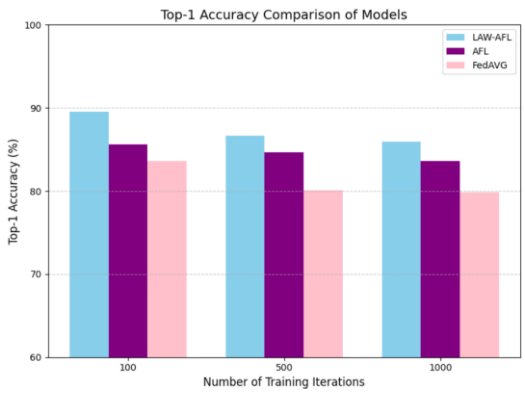
图 10（c）展示了模型在$\alpha = 0.05$数据分布更加异质化的情况下，LAW-AFL 依然占据性能优势，准确率从 100 次迭代的 85%提升至 1000 次

迭代的约 88%。AFL 的表现次优，但与 LAW-AFL 的差距进一步扩大。FedAVG 的准确率显著下降，尤其在 100 次迭代时仅达到 78%，难以应对高异质性数据分布。

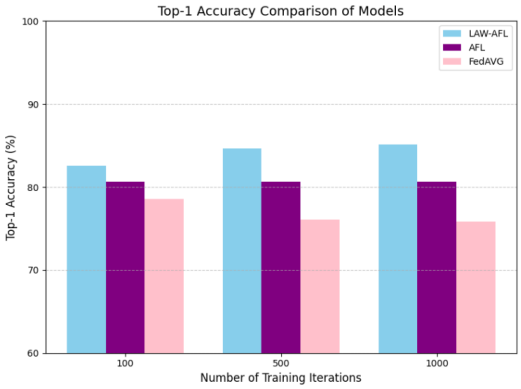
由图 10（d）可知，LAW-AFL 的表现依旧最优，准确率从 100 次迭代的约 78%提升至 1000 次迭代的 86%，增长趋势明显。AFL 的表现与 LAW-AFL 差距显著，最终准确率仅达到 78%。



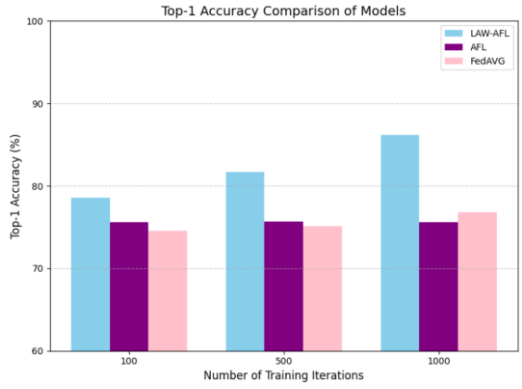
（a）MNIST($\alpha = 0.1$)



（b）CIFAR-100($\alpha = 0.1$)



（c）MNIST($\alpha = 0.05$)



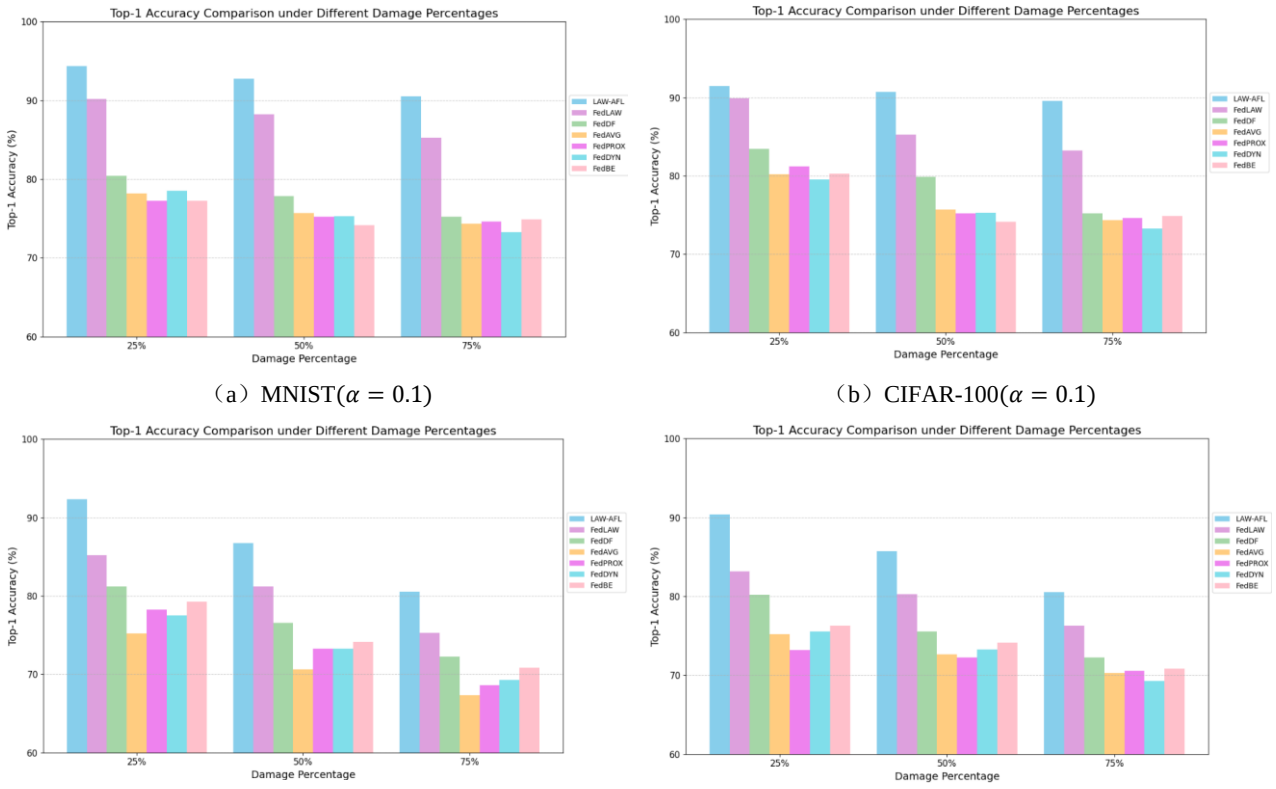
（d）CIFAR-100($\alpha = 0.05$)

图 10　在不同数量客户端下的准确性

4.2.5　不同受损客户端对各模型的影响

在不同破坏比例的客户端（分别为 25%、50%、75%）下，验证 MNIST 和 CIFAR-100 数据集的性能表现，实验结果如图 11 所示，横轴表示客户端破坏的百分比，纵轴为 Top-1 准确率（%）。柱状图中对比了多种联邦学习方法的表现，包括 FedLAW、FedDF、FedAVG 等。从图中可以看出：LAW-AFL 在所有破坏比例下均保持了最高的 Top-1 准确率，无论是 MNIST 还是 CIFAR-100 数据集，均表现出极强的鲁棒性。FedLAW 紧随其后，性能稳定，但在高破坏比例下准确率有所下降。其他方法的表现整体低于 LAW-AFL 和 FedLAW，且随破坏比例增加，性能下降更为明显。FedAVG 的表现最弱，特别是在高破坏比例下准确率显著降低。

总体而言，图 11 反映了 LAW-AFL 在不可靠客户端环境中表现了出色的鲁棒性，证明其适用于高噪声或数据破坏严重的场景。



（a）MNIST($\alpha = 0.1$)　　（b）CIFAR-100($\alpha = 0.1$)

（c）MNIST($\alpha = 0.05$)　　（d）CIFAR-100($\alpha = 0.05$)

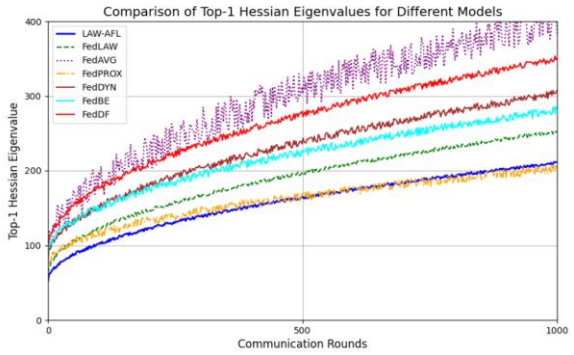
图 11　在不同比例的受损客户端下的性能表现

4.2.6　各模型在不同轮次下的 Top-1 Hessian 特征值对比

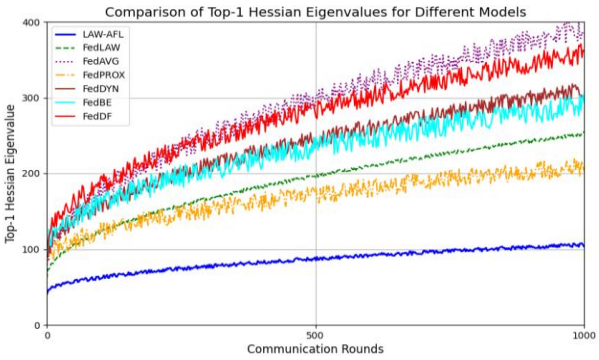
在不同通信轮次下的 Top-1 Hessian 特征值的变化趋势，分为 MNIST 和 CIFAR-100 两个数据集的实验结果，实验结果如图 12 所示，图中对比了七种联邦学习方法，横轴表示通信轮次 Communication Rounds，范围为 0 到 1000。纵轴表示 Hessian 矩阵的 Top-1 特征值，范围为 0 到 400。蓝色曲线为 LAW-AFL 保持最低且平稳，图 12（a）中，LAW-AFL 的解析结构与自适应聚合策略有助于降低震荡，使得特征值保持在 100~200 之间未产生剧烈波动，而 FedAVG、FedPROX 等方法波动较大且上升迅速，表明无法有效应对数据异质性带来的优化震荡。图 12（b）中，数据复杂性显著高于 MNIST，各模型波动的幅度相较于图 12（a）也更加剧烈，LAW-AFL 稳定控制特征值在 150 左右。图 12（c）是在更高的数据异质性，LAW-AFL 相较其他算法仍保持最低且平稳的曲线，说明算法在高异质性下具有稳健性。图 12（d）中 LAW-AFL 仍然保持较高的稳定性在高异质性的环境下，相较于前三个实验结果，LAW-AFL 的特征值呈现上升趋势，但也稳定在 170 左右，其他基线模型呈现明显震荡，最高接近 400。

Top-1 Hessian 特征值增长幅度最小，且波动极小。特征值增长缓慢表明 LAW-AFL 的优化过程更加平滑，对模型参数的更新控制较好，避免了过度震荡。

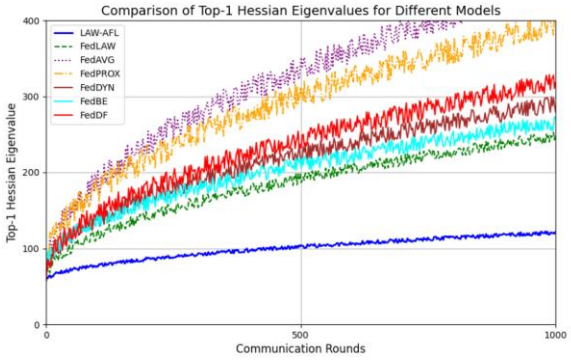
这种稳定性说明 LAW-AFL 更擅长在联邦学习环境中保持模型收敛的稳健性，尤其在复杂的且数据非独立性强的数据集上依然表现出色。



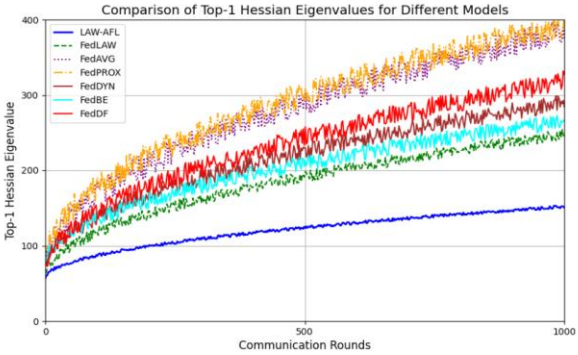
（a）MNIST($\alpha = 0.1$)



（b）CIFAR-100($\alpha = 0.1$)



（c）MNIST($\alpha = 0.05$)



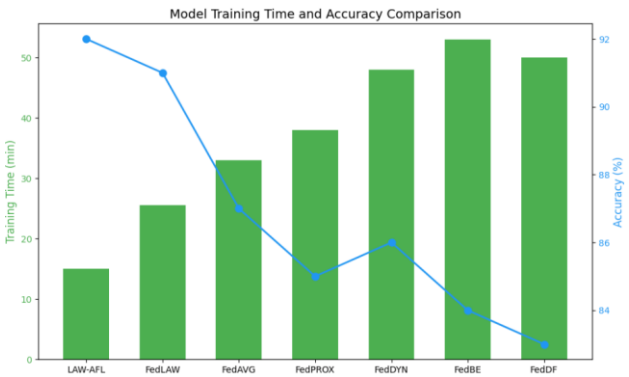
（d）CIFAR-100($\alpha = 0.05$)

图 12　在不同通信轮次下的 Top-1 Hessian 特征值的变化趋势

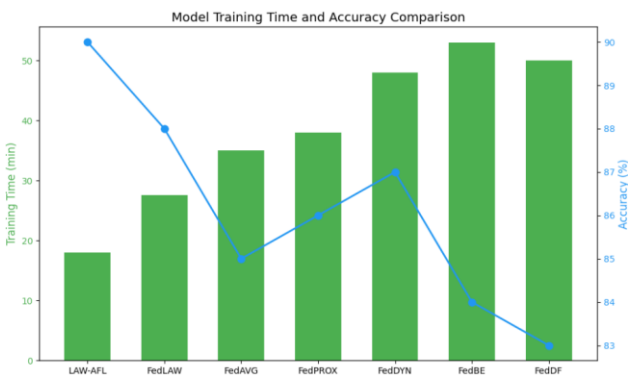
4.2.7　不同模型的时间复杂度与性能分析

在这组实验中，分别对比了 LAW-AFL 模型与其他六种基线模型在 MNIST 和 CIFAR-100 两种数据集下以及不同数据分布参数下的训练时间和模型准确率，结果如图 13 所示。在不同数据集下 LAW-AFL 模型训练 1000 次所需的时间最短，且能达到最高的准确率，因为 LAW-AFL 闭式训练范式，客户端仅需单周期完成本地模型更新。而 FedAVG 等基线方法要求客户端执行多周期迭代，其本地训练时间随 E 线性增长。这证明了 LAW-AFL 模型在通信效率和计算效率是有着显著优势。

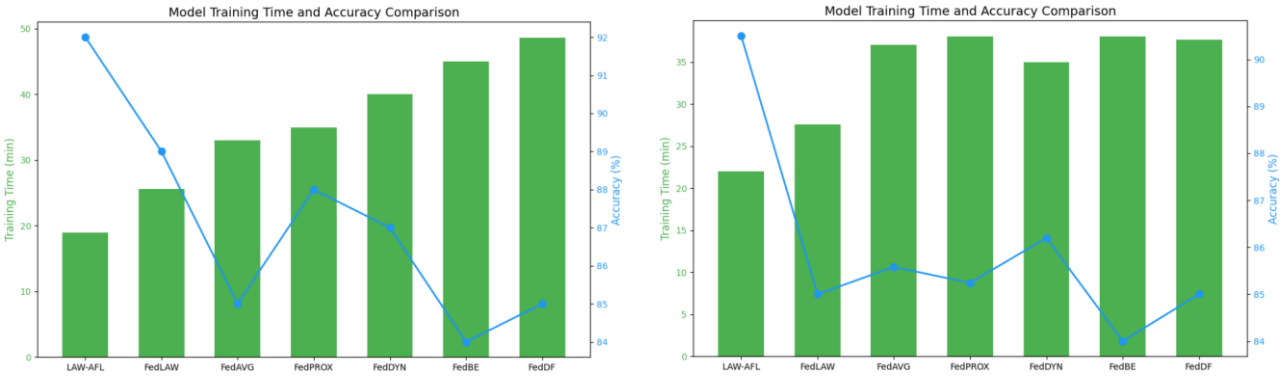
FedLAW 模型训练时间仅次于 LAW-AFL 模型，在数据分布参数更小的环境下，该模型受到的影响较大，特别是 FedDF 模型随着 Non-IID 程度增加而显著下降，这更加表明了 LAW-AFL 模型适合在通信资源有限、数据分布高异质性的实际应用部署。



（a）MNIST($\alpha = 0.1$)



（b）CIFAR-100($\alpha = 0.1$)



（c）MNIST($\alpha = 0.05$)　　（d）CIFAR-100($\alpha = 0.05$)

图 13　1000 次训练迭代下各模型的准确率和时间的对比

4.2.8　通信成本分析

在联邦学习中，通信成本是衡量系统效率的重要指标。为了验证可学习聚合权重在降低通信负担的同时提升模型性能的优势，LAW-AFL 通过代理数据集动态学习聚合权重，并结合自适应收缩因子降低通信负担。设计了一系列通信成本对比实验。实验以 CIFAR-10 数据集($\alpha = 0.1$)为基准，探讨各方法在单轮通信量、全局通信轮次以及总体通信量上的表现，如表 12 所示。

表 12　CIFAR-10($\alpha = 0.1$)设置下通信成本和准确率对比

方法	单轮通信量（MB）	总通信轮次	总通信量（MB）	准确率（%）
FedAVG	18.2	150	2730	78.3
FedPROX	17.8	160	2848	81.5
AFL	12.6	200	2520	76.2
FedLAW	13.5	160	2461	78.4
LAW-AFL	**12.4**	**120**	**1488**	**85.7**

LAW-AFL 的单轮通信量为 12.4MB，相较于 FedAVG 的 18.2MB 减少了约 31.9%。这一结果表明，通过动态调整聚合权重，LAW-AFL 在每次通信时传输的数据量显著减少，有效降低了通信压力。LAW-AFL 仅需 120 轮全局通信，整体通信量仅为 2288MB，相比 FedAVG 的 2730MB 仅为其 54.5%，节省了约 35.5%的通信开销。这意味着在实现更高准确率的同时，LAW-AFL 大幅降低了整个训练过程中的通信资源消耗。通过动态学习和调整权重，LAW-AFL 不仅有效降低了每轮及总通信量，而且显著提高了模型准确率，相较于传统的 FedAVG、FedPROX 及其他基线方法，LAW-AFL 在通信成本和整体训练效率上均具有明显优势。

5　总结

本文针对传统联邦学习框架中因权重归一化和数据异质性导致的过拟合、模型收敛缓慢以及通信成本高昂等挑战，提出了一种创新性算法—LAW-AFL。该算法通过可学习的聚合权重机制和解析性联邦学习范式，有效提升了全局模型的泛化能力，同时优化了训练与通信效率。

LAW-AFL 的核心创新包括以下两个方面：一是基于可学习的聚合权重策略，利用客户端一致性机制增强模型在非独立同分布数据环境下的适应性；二是通过闭形式的单周期训练技术消除了超参数调优的复杂性，从而简化了训练流程，显著提高了训练效率和模型的鲁棒性。LAW-AFL 不仅在理论上提供了一种有效的联邦学习优化方案，在实践中也展现了其在精确性、鲁棒性和效率方面的潜力，为实际应用场景中的联邦学习任务提供了有力支持。

在未来的研究中，可进一步扩展 LAW-AFL 在更多复杂数据场景和任务中的应用，以全面验证其通用性和稳定性。

参考文献

[1] Ma Z,Zhao M,Cai X,et al.Fast-convergent federated learning with class-weighted aggregation.Journal of Systems Architecture, 2021, 117: 102125.

[2] McMahan B., Moore E., Ramage D., et al. Communication-efficient learning of deep networks from decentralized data//Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. Fort Lauderdale, USA, 2017: 1273-1282.

[3] Wang Xin-Ao, Chen Ke, Shou Li-Dan, et al. A High-Efficiency Training Framework for BERT Model Based on Federated Learning. Journal of Software, 2023,51(01): 1–24. (in Chinese)
(王鑫澳,陈珂,寿黎但,等.基于联邦学习的 BERT 模型高效训练框架.软件学报, 2023,51(01): 1-24)

[4] Li T., SahuA K, Talwalkar A,et al. Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine,2024, 37(3): 50-60.

[5] Liu Song, Luo Yang-Yu, Xu Jia-Pei, et al. Low-Cost Federated Learning Based on Lightweight Self-Distillation.ActaElectronic Sinica, 2025, 53(01): 259–269. (in Chinese)
(刘松,罗杨宇,许佳培,等.基于轻量自蒸馏的低成本联邦学习.电子学报,2025,53(01):259-269)

[6] Lin, T., Kong, L., Stich, S. U., and Jaggi, M. Ensemble distillation for robust model fusion in federated learning.Advances in Neural Information Processing Systems,2020, 33:2351–2363.

[7] Lin F P-C,Hosseinalipour S, Michelusi N, et al. Delay-Aware Hierarchical Federated Learning. IEEE Transactions on Cognitive Communications and Networking, 2024, 10(2): 674-688.

[8] He, X., Peng, C., & Tan, W. (2023). Fast and accurate deep leakage from gradients based on wasserstein distance. International Journal of Intelligent Systems, 2023(1), 5510329.

[9] Wang Z, SONG M, ZHANG Z, et al. Beyond inferring class representatives: User-level privacy leakage from fed erated learning.//Proceedings of theIEEE INFOCOM 2019 - IEEE Conference on Computer Communications, Paris, France, 2019: 2512-2520.

[10] Hu X, LI R, WANG L, et al. A Data Sharing Scheme Based on Federated Learning in IoV. IEEE Transactions on Vehicular Technology, 2023, 72(9): 11644-11656.

[11] Liu W, XU X, LI D, et al. Privacy Preservation for Federated Learning With Robust Aggregation in Edge Computing. IEEE Internet Things Journal, 2023, 10(8): 7343-7355.

[12] Li D, LAI J, WANG R, et al. Ubiquitous intelligent federated learning privacy-preserving scheme under edge computing. Future Generation Computer Systems, 2023, 144: 205-218.

[13] Peng Z, XU J, CHU X, et al. VFChain: Enabling Verifiable and Auditable Federated Learning via Blockchain Systems.IEEE Transactions on Network Science and Engineering, 2022, 9(1): 173-186.

[14] Tian Li, Anit Kumar Sahu, Manzil Zaheer, et al. Federated optimization in heterogeneous networks//Proceedings of Machine Learning and Systems Conference,Austin, USA,2020,2: 429–450.

[15] Durmus, Alp Emre, Yue Z, et al, What-mough,P., and Saligrama, V.Federated learning based on dynamic regularization.//Proceedings of theInternational Conference on Learning Representations,Virtual,2021:1-36.

[16] Li Z., Lin T., Shang X, et al. Revisiting Weighted Aggregation in Federated Learning with Neural Networks//Proceedings of the 40th International Conference on Machine Learning.Honolulu, USA,2023, 202:19767-19788.

[17] Guo P, Michael R Lyu, and NE Mastorakis. Pseudoinverse learning algorithm for feedforward neural networks. Advances in Neural Networks and Applications, 2004, 1: 321–326.

[18] Chen S., Billings S. A., Luo W. Orthogonal least squares methods and their application to non-linear system identification. International Journal of Control, 1989, 50(5): 1873-1896.

[19] Kar-Ann Toh. Learning from the kernel and the range space.//Proceedings of the2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS), Singapore,2018: 1–6.

[20] Wang X Z, Zhang T, and Wang R. Noniterative deep learning: Incorporating restricted Boltzmannmachine into multilayer random weight neural networks.IEEE Transactions on Systems, Man and Cybernetics: Systems, 2017, 49(7):1299–1308.

[21] Wang J, Guo P, and Li Y. DensePILAE: a feature reuse pseudoinverse learning algorithm for deep stacked autoencoder. Complex & Intelligent Systems, 2022, 8(3):2039–2049.

[22] Zhuang H, Lin Z, andToh K. Blockwise recursive Moore-Penrose inverse for network learning. IEEE Transactions on Systems, Man and Cybernetics: Systems, 2021, 52(5):3237–3250.

[23] Zhuang H,He R, Tong K, et al. Analytic Federated Learning.arXiv preprint, arXiv:2405.16240,2024.

[24] LiY,Huang C, Wang S, et al. Privacy-Preserving Federated Primal—Dual Learning for Nonconvex and Nonsmooth Problems With Model Sparsification. IEEE Internet of Things Journal, 2024, 11(15): 25853-25866.

[25] LuX, Zheng H, Liu W, et al. POP-FL: Towards Efficient Federated Learning on Edge Using Parallel Over-Parameterization.IEEE Transactions on ServicesComputing, 2024, 17(2): 617-630.

[26] Kwon J, Kim J, Park H, et al. Asam: Adaptive sharpness-aware minimization for scale-invariant learning of deep neural networks//Proceedings of the 38th International Conferenceon Machine Learning,Virtual,2021, 139: 5905–5914.

[27] Xia Y, Yang D, Li W, et al. Auto-fedavg: learnable federated averaging for multi-institutional medicalimage

segmentation.arXiv preprint, arXiv:2104.10195, 2021.

[28] Li S, Zhou T, Tian X, et al. Learning to collabo-rate in decentralized learning of personalized models //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,Location, New Orleans, USA, 2022: 9766-9775.

[29] Charles Z, Garrett Z, Huo Z, et al. On large-cohort training for federated learning. Advances in neural information processing systems,2021, 34:20461–20475.

[30] Tang Ling-Tao, Wang Di, Liu Sheng-Yun. A Data Augmentation Scheme for Federated Learning with Non-IID Data. Journal of Communications, 2023, 44(01): 164–176. (in Chinese)
(汤凌韬,王迪,刘盛云.面向非独立同分布数据的联邦学习数据增强方案.通信学报,2023,44(01):164-176.)

[31] ZhuZ, Shi Y, Fan P, et al, ISFL: Federated Learning for Non-i.i.d. Data With Local Importance Sampling. IEEE Internet of Things Journal, 2024, 11(16): 27448-27462.

[32] Randall E. Cline. Representations for the generalized inverse of a partitioned matrix. Journal ofthe Society for Industrial and Applied Mathematics,1964, 12(3):588–600.

[33] Wang H, Yurochkin M, Sun Y,et al. Federated learning with matched averaging//Proceedings of International Conference onLearning Representations, Virtual, 2020.

[34] Chen, H. and Chao, W. Fedbe: Making bayesian model ensemble applicable to federated learning//Proceedingsof the 9th International Conference on Learning Representations, ICLR2021, Austria, 2021: 3-7.

[35] Jiang Wei-Jin, Du Xi-Chen, Jiang Yi-Rong, et al. A swarm intelligence perception algorithm for environmental monitoring based on adaptive federated learning Electronic Journal, 2025, 53 (03): 821-835. (in Chinese)
(蒋伟进,杜熙晨,蒋意容,等.基于自适应联邦学习的环境监测群智感知算法.电子学报,2025,53(03):821-835.)

JIANG Wei-Jin,Ph.D., Professor. His main research interests include edge computing, social computing, and cyberspace security.

CUI Xin-Yu, M.S. candidate,Her main research interests include federated learning and privacy protection.

Liu Zhi-Hua,M.S. candidate, His main research interests include federated learning and privacy protection.

CHEN Shen-You, M.S. candidate, His main research interests include federated learning and privacy protection.

HU Jia-Long, M.S. candidate, His main research interests include dynamic pricing of data and federated learning.

Background

This study addresses challenges within federated learning (FL), particularly in scenarios involving non-independent and identically distributed (Non-IID) data and large-scale client participation. FL performance and communication efficiency often degrade significantly under these conditions. Current FL methods generally adopt fixed weight normalization aggregation strategies, which fail to adequately account for data heterogeneity among clients. This limitation leads to a decline in the global model’s generalization performance and slower convergence, limiting FL's applicability in complex real-world scenarios.

As client participation and data volume grow, the communication overhead in FL increases, exacerbating training time and resource demands. Additionally, existing methods rely heavily on extensive hyperparameter tuning, which is time-consuming and requires significant manual intervention, making practical deployment more difficult. Given the diverse nature of data and tasks, achieving efficient convergence while improving global model performance remains an urgent challenge. The typically Non-IID nature of client data further complicates existing aggregation strategies, impairing model generalization and stability.

To address these issues, this paper proposes an adaptive federated learning method based on

learnable aggregation weights. By incorporating a learnable aggregation weight strategy, the proposed method enhances the model's adaptability to complex tasks while preserving data privacy, promoting the integration of privacy-preserving techniques with data-driven technologies. The method's improvement of aggregation strategies and optimization of the training paradigm offer a novel approach to tackling challenges related to data heterogeneity and large-scale client participation. This approach significantly enhances the accuracy and robustness of the global model.

By simplifying the hyperparameter tuning process and reducing communication overhead, the method supports the efficient deployment of FL and facilitates practical applications in fields like healthcare and finance.

The significance of this project lies in providing theoretical and technical support for FL's application in privacy-sensitive domains, such as healthcare and finance, while laying the foundation for efficient distributed learning in large-scale Non-IID data scenarios. The research outcomes not only improve FL algorithm performance but also reduce communication costs in real-world applications, offering solutions for privacy-preserving computation and collaborative optimization in an intelligent society.

The research group has previously made significant strides in FL, including optimizing aggregation strategies, designing personalized FL frameworks, and exploring multi-model collaborative optimization, which provide a solid foundation for this study.