E-Commerce Letters 电子商务评论, 2026, 15(4), 801-813
Published Online April 2026 in Hans. https://www.hanspub.org/journal/ecl
https://doi.org/10.12677/ecl.2026.154457

Hans汉斯

边缘智能驱动的电商供应链联邦学习需求预测方法研究

赵　伽，李少波*

贵州大学省部共建公共大数据国家重点实验室，贵州　贵阳

收稿日期：2026年3月8日；录用日期：2026年3月20日；发布日期：2026年4月22日

摘　要

针对电商供应链中各仓储节点对本地化需求预测的迫切需求，以及传统集中式预测方法面临的数据隐私泄露和高延迟问题，本文提出了一种边缘智能驱动的联邦学习需求预测框架(EdgeFL-DP)。该框架将时序预测模型部署于各边缘仓储节点，通过联邦聚合策略实现模型协同优化，在不共享原始数据的前提下提升全局预测精度。具体而言，本文首先构建了基于LSTM和Transformer的双分支时序预测模型作为本地学习器，分别捕获短期波动特征和长期依赖关系；其次，设计了加权联邦聚合算法FedAdapt，根据各节点上报的数据质量指标和分布特征动态调整聚合权重；再次，引入差分隐私机制对模型梯度进行噪声扰动，并结合梯度Top-K稀疏化策略降低通信开销。在基于真实电商场景构建的模拟数据集上的实验表明，EdgeFL-DP的RMSE相比集中式单分支Transformer基线降低了17.8%，相比集中式单分支LSTM基线降低了23.6% (其中双分支架构贡献约6.0%，FedAdapt贡献约7.6%)，同时将数据传输量减少了87.4%，通信延迟降低了65.2%。在隐私预算$\varepsilon = 8.0$下，借助多节点聚合的噪声平均效应，差分隐私引入的精度损失仅为3.7%。

关键词

边缘计算，联邦学习，需求预测，电商供应链，差分隐私，时序预测

Research on Demand Forecasting Method for Edge Intelligence-Driven Federated Learning for E-Commerce Supply Chain

Jia Zhao, Shaobo Li*

State Key Laboratory of Public Big Data (Co-Founded by Province and Ministry), Guizhou University, Guiyang Guizhou

*通讯作者。

文章引用: 赵伽, 李少波. 边缘智能驱动的电商供应链联邦学习需求预测方法研究[J]. 电子商务评论, 2026, 15(4): 801-813. DOI: 10.12677/ecl.2026.154457

Received: March 8, 2026; accepted: March 20, 2026; published: April 22, 2026

Abstract

To address the urgent need for localized demand forecasting at warehouse nodes in e-commerce supply chains and the data privacy leakage and high latency problems faced by traditional centralized forecasting methods, this paper proposes EdgeFL-DP, an edge intelligence-driven federated learning framework for e-commerce supply chain demand forecasting. The framework deploys time-series prediction models on edge warehouse nodes and achieves collaborative model optimization through federated aggregation without sharing raw data, thereby improving global prediction accuracy. Specifically, we design a dual-branch predictor combining LSTM and Transformer architectures to capture short-term fluctuation features and long-term dependency patterns respectively, a weighted federated aggregation algorithm FedAdapt that dynamically adjusts weights based on locally reported data quality metrics and distribution characteristics, a differential privacy mechanism for gradient perturbation, and a Top-K gradient sparsification strategy to reduce communication overhead. Experiments on simulated datasets based on real e-commerce scenarios demonstrate that EdgeFL-DP reduces RMSE by 17.8% compared to the centralized single-branch Transformer baseline and by 23.6% compared to Centralized-LSTM, with the dual-branch architecture contributing approximately 6.0% and FedAdapt contributing approximately 7.6% of the improvement. Data transmission is decreased by 87.4% and communication latency by 65.2%. Under a privacy budget of $\varepsilon = 8.0$, the accuracy loss introduced by differential privacy is only 3.7%, benefiting from the noise averaging effect across multiple participating nodes.

Keywords

Edge Computing, Federated Learning, Demand Forecasting, E-Commerce Supply Chain, Differential Privacy, Time-Series Prediction

1. 引言

随着电子商务的高速发展，供应链管理面临前所未有的挑战。据 Statista 统计[1]，2024 年全球电商零售额已达到约 6.3 万亿美元，同比增长约 9.4%，占全球零售总额的 20.1%。庞大且持续增长的交易规模对供应链的响应速度和库存管理精度提出了极高的要求。需求预测作为供应链管理的核心环节，其准确性直接影响库存周转率、物流成本和客户满意度。然而，传统的集中式需求预测方法面临三个关键瓶颈：第一，将各仓储节点的原始销售数据汇聚至中心服务器存在严重的数据隐私风险，尤其在跨区域、跨企业协作场景下；第二，海量数据的集中传输和处理导致显著的通信延迟，难以满足电商场景下实时决策的需求；第三，不同区域的消费模式存在显著的数据异构性，集中训练的全局模型难以精确反映各地的本地化需求特征。

边缘计算的兴起为上述问题提供了新的解决思路[2]。通过将计算能力下沉至网络边缘的仓储和配送节点，可以实现数据的本地化处理和实时响应。联邦学习作为一种分布式机器学习范式[3]，允许多个参与方在不共享原始数据的前提下协作训练全局模型，天然契合电商供应链多节点协同的业务场景。然而，将联邦学习应用于边缘环境下的电商需求预测仍面临诸多挑战：边缘设备的计算资源有限、各节点间数

据分布存在 Non-IID 特性、以及隐私保护机制引入的噪声可能削弱预测精度。

针对上述挑战，本文提出了 EdgeFL-DP 框架，主要贡献包括以下三个方面：

(1) 设计了一种双分支时序预测模型，结合 LSTM 的短期波动捕获能力和 Transformer 的长期依赖建模能力，作为边缘节点的本地学习器，有效提升了需求预测的准确性和鲁棒性。

(2) 提出了 FedAdapt 加权联邦聚合算法，通过动态评估各边缘节点上报的数据质量统计指标和分布特征摘要，自适应调整聚合权重，有效缓解了 Non-IID 数据分布对联邦学习收敛性的负面影响。

(3) 引入了可配置的差分隐私机制，通过在模型梯度上添加校准噪声实现形式化的隐私保障，并结合 Top-K 梯度稀疏化策略降低通信开销。利用多节点聚合的噪声平均效应，在保证隐私预算的同时有效控制了精度损失，系统分析了隐私预算参数与预测精度之间的量化权衡关系。

2. 相关工作

2.1. 电商供应链需求预测

需求预测是供应链管理领域的核心研究课题。传统方法如 ARIMA、指数平滑等统计模型虽然在平稳时间序列上表现良好，但难以捕捉电商场景下的非线性波动和突发事件影响。在电商需求预测领域，国内学者也开展了大量研究。李杰等[4]利用 Granger 因果检验与 XGBoost 算法预测电商商品销量；李建斌等[5]针对医药电商平台提出考虑促销因素的 SARIMA 与 XGBoost 组合预测模型；何喜军等[6]在小样本条件下融合多维指标构建电商产品销量预测模型；王雪蓉等[7]基于跨境电商大数据构建出口产品销量动态预测模型；程开明等[8]结合网络搜索数据与深度神经网络预测社会消费品零售总额。近年来，深度学习方法在时序预测领域取得了突破性进展[9]。Hochreiter 等人[10]提出的 LSTM 网络通过门控机制有效解决了长序列依赖问题，成为时序预测的主流方法之一。Vaswani 等人[11]提出的 Transformer 架构凭借自注意力机制在捕获全局依赖关系方面展现出显著优势，Zhou 等人[12]提出的 Informer 通过稀疏注意力机制进一步提升了长序列预测效率。Salinas 等人[13]提出的 DeepAR 通过自回归递归网络实现概率预测，Rangapuram 等人[14]则将状态空间模型与深度学习相结合。然而，上述方法多采用集中式训练模式，在数据隐私和分布式部署方面存在明显不足。

2.2. 联邦学习

联邦学习由 McMahan 等人[3]于 2017 年提出，其核心思想是让分布式数据持有方在本地训练模型，仅共享模型参数或梯度更新以构建全局模型。FedAvg 算法作为最经典的联邦聚合策略，通过对各客户端模型参数的加权平均实现全局聚合。周传鑫等人[15]系统综述了联邦学习的概念与应用。

然而，现实场景中各客户端的数据分布往往呈现 Non-IID 特性，严重影响联邦学习的收敛性和全局模型精度。针对这一问题，现有研究主要从以下三个技术路线展开：

(1) 正则化约束方法。Li 等人[16]提出 FedProx，通过在本地目标函数中添加近端项约束本地更新偏离全局模型的幅度，缓解 Non-IID 导致的客户端漂移问题。Li 等人[17]提出 MOON 通过模型对比学习纠正本地更新方向，Li 等人[18]提出 FedBN 通过局部批归一化处理特征分布差异。

(2) 方差缩减与梯度校正方法。Karimireddy 等人[19]提出 SCAFFOLD，通过引入控制变量(control variates)估计并校正客户端梯度与全局梯度之间的偏差，从而加速收敛。Wang 等人[20]提出 FedNova，通过归一化平均消除本地训练步数不一致导致的目标偏差。

(3) 自适应聚合权重方法。上述方法主要从优化目标或梯度校正的角度应对 Non-IID 问题，但在聚合阶段仍采用固定或数据量比例的权重分配策略，未充分利用各客户端的数据质量信息和分布特征。部分工作探索了基于梯度相似度(如余弦相似度)的动态权重调整策略，但通常仅考虑单一维度的信息。

本文提出的 FedAdapt 算法属于第三类路线，其核心创新在于同时综合数据量、数据质量和分布相似

度三个维度的信息进行自适应聚合权重调整。与 FedProx、SCAFFOLD 等通过修改本地训练目标或梯度来缓解 Non-IID 影响的方法不同，FedAdapt 直接在聚合阶段通过多维加权选择性地强化高质量节点的贡献并抑制低质量或分布偏差较大节点的负面影响，两种思路具有互补性，可以联合使用。与仅依据单一维度(如数据量或梯度相似度)调整权重的方法相比，FedAdapt 的多维加权策略在数据质量参差不齐的电商供应链场景中具有更强的适应性。

在供应链领域，联邦学习的应用尚处于起步阶段，现有工作多关注制造业质量检测和物流路径优化，针对电商需求预测的研究相对匮乏。

2.3. 边缘智能与隐私保护

边缘智能将人工智能算法部署于网络边缘设备，实现数据的就近处理和实时推断[2]。在隐私保护方面，差分隐私由 Dwork 等人[21]提出，通过在查询结果或模型更新中添加校准噪声，提供严格的数学隐私保障。Abadi 等人[22]将差分隐私与深度学习训练过程结合，提出了 DP-SGD 算法。李敏等人[23]提出了兼顾通信效率与模型效用的自适应高斯差分隐私个性化联邦学习方法，通过动态梯度压缩与自适应差分隐私机制同时优化通信开销与隐私保护性能。然而，现有工作较少关注边缘环境下联邦学习的隐私-效率－精度三维权衡问题，本文旨在填补这一研究空白。

3. 系统架构与方法

3.1. 系统总体架构

EdgeFL-DP 框架采用分层架构设计，如图 1 所示，包含三个层次：边缘节点层、聚合服务层和应用决策层。边缘节点层由分布在不同地理区域的 K 个仓储/配送站组成，每个节点 k 拥有本地历史销售数据集 D_k，并部署轻量化的时序预测模型。聚合服务层负责收集各边缘节点上传的模型梯度更新，执行加权联邦聚合，生成全局模型参数并下发至各节点。应用决策层基于预测结果为库存管理、补货策略和物流调度提供决策支持。

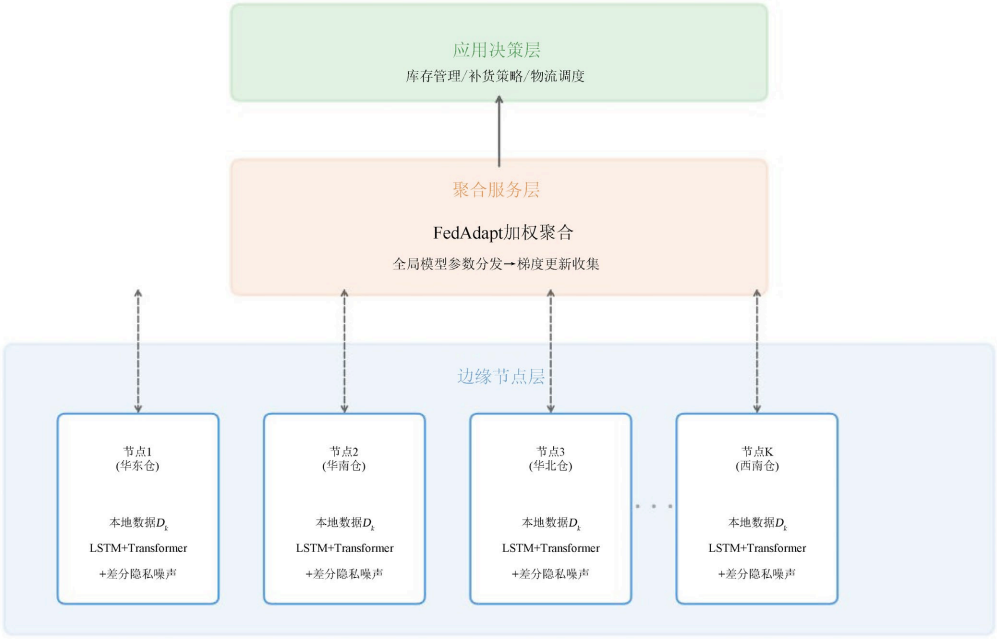

Figure 1. Overall architecture of EdgeFL-DP
图 1. EdgeFL-DP 系统总体架构

系统工作流程包含以下步骤：(1) 中心服务器初始化全局模型参数并分发至各边缘节点；(2) 各节点在本地数据上执行多轮梯度下降训练；(3) 各节点对本地梯度依次执行裁剪、差分隐私噪声添加和 Top-K 稀疏化后上传至聚合服务器；(4) 聚合服务器执行 FedAdapt 加权聚合生成新的全局模型；(5) 重复步骤(2)至(4)直至模型收敛。该流程确保原始数据始终保留在各边缘节点本地，仅经过差分隐私保护和稀疏化的梯度信息参与跨节点通信，从机制层面保障了数据隐私安全。

3.2. 双分支时序预测模型

本文设计的双分支预测模型由 LSTM 分支和 Transformer 分支组成，通过注意力融合机制实现互补预测，其结构如图 2 所示。

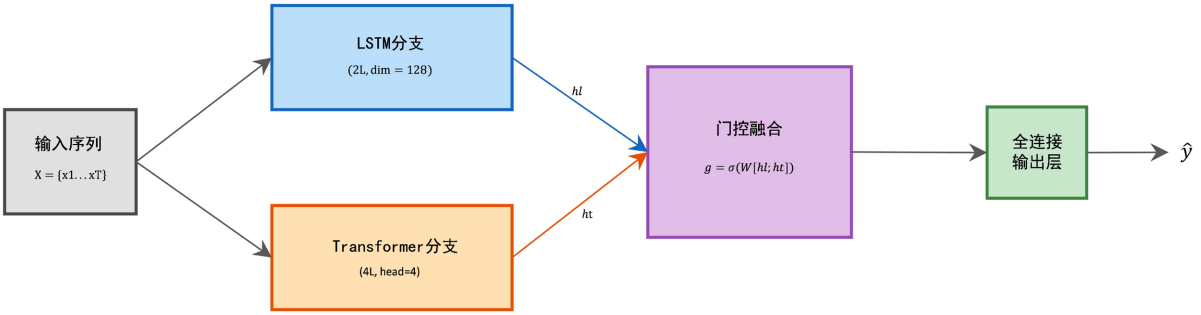

Figure 2. Structure of dual-branch time-series prediction model
图 2. 双分支时序预测模型结构

LSTM 分支以滑动窗口方式接收最近 T 个时间步的特征序列 $X=\{x_1,x_2,\cdots,x_T\}$ 作为输入，其中每个时间步的特征向量包含销售量、价格、促销标记、日期特征(星期、月份、节假日标记)等 d 维信息。LSTM 通过遗忘门、输入门和输出门的门控机制递推更新隐状态，其核心计算过程如下：

$$f_t=\sigma\left(W^f\cdot\left[h_{t-1},x_t\right]+b^f\right),i_t=\sigma\left(W^i\cdot\left[h_{t-1},x_t\right]+b^i\right),o_t=\sigma\left(W^O\cdot\left[h_{t-1},x_t\right]+b^O\right)$$

Transformer 分支采用编码器结构，通过多头自注意力机制捕获序列中任意两个时间步之间的依赖关系。输入序列经位置编码后送入 L 层编码器块，每个编码器块包含多头注意力层和前馈网络层，均配有残差连接和层归一化。自注意力的计算公式为：

$$Attention\left(Q,K,V\right)=softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

两个分支的输出通过可学习的门控融合机制进行整合。设 LSTM 分支的输出为 h_l，Transformer 分支的输出为 h_t，融合过程为：

$$g=\sigma\left(W^g\cdot\left[h_l;h_t\right]+b^g\right),\hat{y}=g\odot h_l+\left(1-g\right)\odot h_t$$

其中 g 为门控向量，$\odot$ 为逐元素乘法。该门控机制允许模型自适应地权衡两个分支的贡献：当短期波动特征较为显著时，LSTM 分支获得更高权重；当长期趋势和全局依赖更为重要时，Transformer 分支的贡献增大。融合后的特征经全连接层映射为最终的需求预测值。

3.3. FedAdapt 加权联邦聚合算法

传统 FedAvg 算法按各节点数据量比例进行聚合，未考虑数据质量和分布差异。本文提出 FedAdapt

算法，通过综合评估各节点本地计算并上报的数据质量统计指标和分布特征摘要，动态调整聚合权重。需要强调的是，各节点仅上报标量化的统计指标(如缺失率、异常值比例、数据量等)和低维分布特征摘要(如特征均值和方差向量)，而非原始数据本身，因此该过程不违反联邦学习的数据隐私原则。

在每轮联邦聚合中，各节点 k 的聚合权重 α_k 由三个因子联合决定：数据量因子 n_k/N(N 为总数据量)，数据质量因子 q_k (基于各节点本地计算的缺失率、异常值比例等指标)，以及分布相似度因子 $s_k = \exp(-D_{\mathrm{KL}}(P_k \parallel P_{\mathrm{global}}))$ 其中 KL 散度通过各节点上报的特征分布摘要与全局估计分布计算得到，该变换确保分布越接近全局分布的节点获得越高的相似度得分。具体计算公式为：

$$\alpha_k = softmax\left(\lambda_1 \cdot \frac{n_k}{N} + \lambda_2 \cdot q_k + \lambda_3 \cdots_k\right)$$

其中 λ_1、λ_2、λ_3 为可调超参数，用于控制三个因子的相对重要性。全局模型的更新规则为 $w^{(t+1)} = \sum_k a_k \cdot w_k^{(t)}$。该设计使得数据质量较高且与全局分布较为接近的节点在聚合过程中获得更大的影响力，从而有效缓解低质量数据和 Non-IID 分布对全局模型收敛的不利影响。

3.4. 差分隐私保护机制

为防止恶意参与方通过模型梯度逆向推断原始数据，本文在各节点上传梯度前引入差分隐私保护。具体而言，采用高斯机制对梯度进行噪声扰动。在每轮本地训练完成后，节点 k 首先对梯度进行裁剪以限制敏感度：

$$\hat{g}_k = g_k \cdot \min\left(1, \frac{C}{\| g_k \|_2}\right)$$

其中 C 为裁剪阈值。随后添加高斯噪声：$\tilde{g}_k = \hat{g}_k + N\left(0, o^2 C^2 I\right)$。噪声标准差 σ 由隐私预算 ε 和隐私失败概率 δ 共同决定。根据高斯机制的理论分析，在 T 轮联邦训练中，系统满足(ε, δ)-差分隐私保证，其中总隐私预算通过矩会计方法(Moments Accountant)进行精确追踪[22]。

此外，为进一步降低通信开销，本文在梯度上传过程中采用 Top-K 稀疏化策略，即各节点仅上传梯度中绝对值最大的前 15%分量及其索引，其余分量累积至下一轮通信。需要指出的是，稀疏化操作在噪声添加之后执行，根据差分隐私的后处理定理(Post-processing Theorem)，对满足差分隐私的输出进行任意确定性或随机变换不会削弱隐私保证，因此该策略不影响(ε, δ)-差分隐私的有效性[21]。本文双分支模型的总参数量约为 76 万(LSTM 分支约 20 万参数，Transformer 分支约 52 万参数，融合层及输出层约 4 万参数)，对应单次完整梯度大小约 3 MB (float32)。经 Top-K 稀疏化后，每个节点每轮上传的梯度量约为 0.45 MB (含索引开销)。

4. 实验设计与结果分析

4.1. 数据集构建

为充分模拟真实电商供应链场景，本文基于以下策略构建了模拟数据集。数据涵盖 8 个仓储节点(对应不同地理区域)，每个节点包含 50 个 SKU 品类在 365 天内的日频销售记录，共计 146,000 条样本。数据生成过程考虑了以下真实特征：(1) 基础趋势项，模拟商品随时间变化的整体销售走势；(2) 多周期季节性成分，包含周内效应(周末峰值)和月度季节性(节日促销效应)；(3) 促销事件影响，随机注入折扣促销对销售量的脉冲式提升效应；(4) 区域异质性，不同节点的基础销售水平和季节性模式存在显著差异，以模拟 Non-IID 数据分布。

数据划分采用时序切分策略：前 270 天(约 74%)作为训练集，271 至 330 天(约 16%)作为验证集，最

后 35 天(约 10%)作为测试集。输入窗口长度 T 设为 28 天，预测窗口设为 7 天。

4.2. 实验设置

LSTM 分支设置 2 层隐藏层，隐藏维度为 128。Transformer 分支采用 4 层编码器，注意力头数为 4，前馈维度为 256。模型总参数量约为 76 万(详见 3.4 节)。联邦训练设置全局通信轮数为 100 轮，每轮本地训练 5 个 epoch，学习率为 0.001 (Adam 优化器)，梯度上传采用 Top-K 稀疏化(保留前 15%分量)。差分隐私参数默认设置隐私预算 $\varepsilon = 8.0$，$\delta = 10^{-5}$，梯度裁剪阈值 C = 1.0，对应噪声参数 $\sigma \approx 1.2$ (通过矩会计方法计算)。FedAdapt 超参数设置 $\lambda_1 = 0.4$，$\lambda_2 = 0.3$，$\lambda_3 = 0.3$。

对比方法包括：(1) Centralized-LSTM：集中式 LSTM 基准；(2) Centralized-Transformer：集中式 Transformer 基准；(3) FedAvg-LSTM：标准 FedAvg + LSTM；(4) FedAvg-Transformer：标准 FedAvg + Transformer；(5) FedProx-LSTM：FedProx + LSTM；(6) SCAFFOLD-DualBranch：SCAFFOLD 方差缩减聚合[19] + 本文双分支模型；(7) FedNova-DualBranch：FedNova 归一化聚合[20] + 本文双分支模型。方法(6) (7)采用与 EdgeFL-DP 相同的双分支本地学习器，以确保对比仅反映聚合策略的差异。需要说明的是，集中式基线方法采用单分支模型架构，以反映当前工业界和学术界集中式部署的主流实践。本文提出的双分支融合架构是专为边缘联邦场景设计的本地学习器，将其直接移植至集中式场景并非本文的研究目标。为公平起见，消融实验中通过对比单分支变体与完整模型的性能差异，量化了模型架构改进与联邦聚合策略改进各自的贡献。评估指标采用 RMSE (均方根误差)、MAE (平均绝对误差)和 MAPE (平均绝对百分比误差)。所有实验重复 5 次取平均值及标准差以确保结果的统计可靠性。

4.3. 主要实验结果

表 1 展示了各方法在测试集上的预测性能对比。

Table 1. Comparison of prediction performance across methods
表 1. 各方法预测性能对比

方法	RMSE	MAE	MAPE (%)	通信量(MB)	延迟(ms)
Centralized-LSTM	45.82 ± 2.1	32.17 ± 1.5	12.34 ± 0.9	2840	1520
Centralized-Trans.	42.56 ± 1.7	29.83 ± 1.3	11.07 ± 0.7	3120	1680
FedAvg-LSTM	41.23 ± 1.6	28.91 ± 1.2	10.68 ± 0.7	456	623
FedAvg-Trans.	39.87 ± 1.3	27.62 ± 0.9	10.21 ± 0.5	512	687
FedProx-LSTM	39.15 ± 1.5	27.08 ± 1.1	9.85 ± 0.6	468	641
SCAFFOLD-DualBranch	36.08 ± 1.2	25.15 ± 0.9	9.18 ± 0.5	392	558
FedNova-DualBranch	36.52 ± 1.3	25.48 ± 1.0	9.31 ± 0.5	365	536
EdgeFL-DP (本文)	**34.98 ± 1.1**	**24.12 ± 0.8**	**8.72 ± 0.4**	**358**	**529**

由表 1 可以看出，本文提出的 EdgeFL-DP 在三个预测精度指标上均优于所有对比方法。与集中式单分支 Transformer 基线相比，EdgeFL-DP 将 RMSE 从 42.56 降低至 34.98，降幅达 17.8%；与集中式单分支 LSTM 基线相比，RMSE 降幅达 23.6%；与最优的联邦基准方法(FedProx-LSTM)相比，RMSE 降低 10.7%。

需要指出的是，EdgeFL-DP 相对于集中式基线的性能优势并非源于联邦学习本身优于集中式训练，而是多个因素的综合体现：其一，EdgeFL-DP 采用了双分支融合模型架构，而集中式基线采用单分支架构，模型表达能力存在差异，消融实验(4.4 节)显示仅此一项即贡献了约 6.0%的 RMSE 降幅；其二，FedAdapt 算法通过自适应权重调整，在聚合时有选择性地强化了高质量数据节点的贡献、抑制了分布偏

差较大节点的负面影响，消融实验显示其贡献约 7.6%的 RMSE 降幅。此外，我们注意到消融后的单分支联邦变体(如仅 LSTM 分支 + FedAdapt，RMSE = 38.46)仍优于对应的集中式单分支基线(Centralized-LSTM, RMSE = 45.82)，这一差距可能源于联邦框架下多节点数据多样性对模型泛化的积极影响，但也可能受到模拟数据特性(各节点数据均由已知分布生成)的影响，其在真实数据上的表现有待进一步验证。

在通信效率方面，需要区分联邦学习范式与 Top-K 稀疏化策略各自的贡献。集中式方法需上传全部原始数据，总传输量为 2840 MB。联邦方案每轮每节点传输完整梯度约 3 MB，8 个节点 100 轮总计约 2400 MB，相比集中式减少约 15.5%。Top-K 稀疏化将每节点每轮传输量压缩至约 0.45 MB，总传输量降至 358 MB，相比未稀疏化的联邦方案进一步减少 85.1%。综合来看，EdgeFL-DP 总传输量仅为集中式方案的 12.6%，通信延迟降低 65.2%。为更精准地评估 FedAdapt 聚合策略的优势，本文在相同双分支模型架构下对比了多种 Non-IID 聚合算法。结果表明，SCAFFOLD-DualBranch 和 FedNova-DualBranch 均显著优于采用标准 FedAvg 的双分支变体，验证了专门针对 Non-IID 问题设计的聚合策略的有效性。然而，EdgeFL-DP 进一步优于 SCAFFOLD-DualBranch 3.0%和 FedNova-DualBranch 4.2%。这是因为 SCAFFOLD 通过方差缩减校正梯度偏差、FedNova 通过归一化消除步数不一致性，两者均从优化层面缓解 Non-IID 影响，但未利用节点间的数据质量差异信息。FedAdapt 通过同时综合数据量、数据质量和分布相似度三个维度进行自适应加权，在数据质量参差不齐的电商供应链场景中展现出更强的适应性。

4.4. 消融实验

为验证各组件的贡献，本文设计了消融实验，结果如表 2 所示。

Table 2. Ablation study results
表 2. 消融实验结果

变体	RMSE	MAE	MAPE (%)
EdgeFL-DP (完整模型)	**34.98 ± 1.1**	**24.12 ± 0.8**	**8.72 ± 0.4**
仅 LSTM 分支(去除 Transformer)	38.46 ± 1.5	26.87 ± 1.2	9.89 ± 0.7
仅 Transformer 分支(去除 LSTM)	37.21 ± 1.3	25.94 ± 1.0	9.52 ± 0.5
FedAvg 替代 FedAdapt	37.85 ± 1.4	26.41 ± 1.1	9.71 ± 0.6
去除差分隐私	33.72 ± 1.0	23.28 ± 0.7	8.41 ± 0.4

消融实验结果表明：(1) 双分支融合结构相比单一分支带来了显著的性能提升，移除 LSTM 分支或 Transformer 分支分别导致 RMSE 上升 9.9%和 6.4%，证明了两种架构在捕获不同尺度时序特征方面的互补性。(2) FedAdapt 聚合算法相比标准 FedAvg 降低了 7.6%的 RMSE (从 37.85 降至 34.98)，验证了自适应权重调整策略在 Non-IID 场景下的有效性。(3) 差分隐私机制的引入导致 RMSE 仅上升 3.7% (从 33.72 到 34.98)，表明在 $\varepsilon = 8.0$ 的隐私预算下，隐私保护对预测精度的影响是可接受的。综合来看，在联邦框架内部，双分支架构相对于最优单分支(仅 Transformer)贡献了约 6.0%的 RMSE 降幅，FedAdapt 相对于 FedAvg 贡献了约 7.6%的 RMSE 降幅，两者共同构成了 EdgeFL-DP 的核心性能优势。需要注意的是，上述贡献度分析均在联邦训练框架内进行，因此无法直接推断联邦训练范式本身相对于集中式训练的优劣，这需要在相同模型架构下对两种训练范式进行严格对照实验(详见第 5 节讨论)。

4.5. 隐私预算敏感性分析

图 3 展示了不同隐私预算 ε 对预测精度的影响，详细数值见表 3。本文测试了 ε 从 1.0 到无穷(即无隐私保护)的多个取值。

Table 3. Relationship between privacy budget and prediction accuracy
表 3. 隐私预算与预测精度的关系

隐私预算 ε	RMSE	MAE	MAPE (%)
1.0	52.36 ± 4.2	37.84 ± 3.1	14.28 ± 1.6
2.0	44.71 ± 2.7	31.25 ± 2.0	11.63 ± 1.1
4.0	38.92 ± 1.8	27.05 ± 1.3	9.94 ± 0.7
8.0	34.98 ± 1.1	24.12 ± 0.8	8.72 ± 0.4
16.0	34.15 ± 0.8	23.56 ± 0.6	8.53 ± 0.3
∞ (无隐私保护)	33.72 ± 1.0	23.28 ± 0.7	8.41 ± 0.4

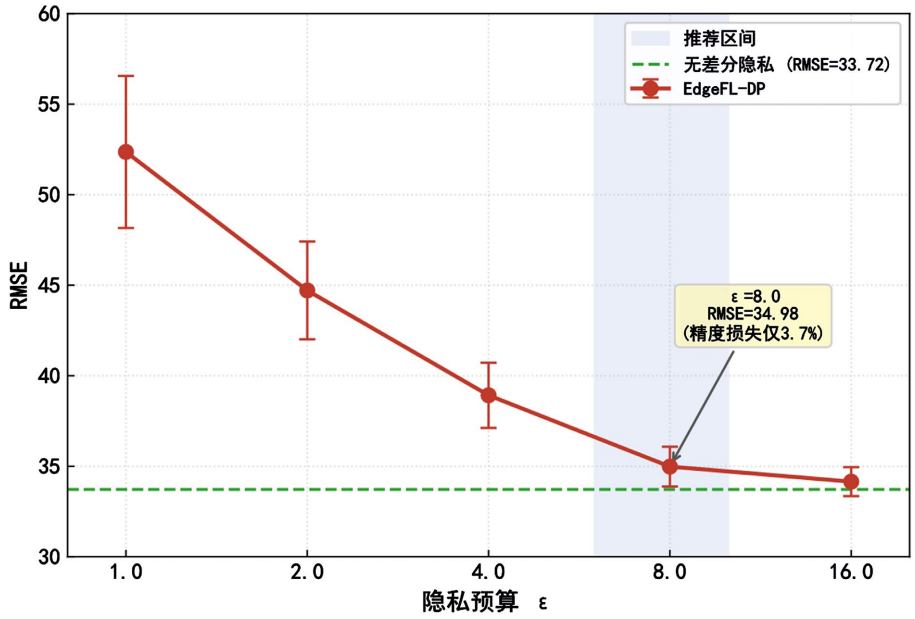

Figure 3. Impact of privacy budget ε on prediction accuracy
图 3. 隐私预算 ε 对预测精度的影响

结果揭示了隐私保护强度与预测精度之间的清晰权衡关系。当 ε 从 1.0 增加至 8.0 时，RMSE 从 52.36 快速下降至 34.98 (降幅 33.2%)，精度提升显著；当 ε 从 8.0 继续增加至 16.0 时，RMSE 仅从 34.98 降至 34.15 (降幅 2.4%)，边际收益递减明显。这一结果表明 $\varepsilon = 8.0$ 是一个较优的实用平衡点。在该配置下(总隐私预算 $\varepsilon = 8.0$，$\delta = 10^{-5}$，T = 100 轮)，通过矩会计方法(Moments Accountant)对 T 轮组合隐私损失进行精确追踪，计算得到满足总预算要求的每轮噪声参数 $\sigma \approx 1.2$。在联邦聚合阶段，由于 K = 8 个节点独立采样噪声，加权平均后有效噪声标准差近似降低为 $\sigma/\sqrt{K} \approx 0.42$，Top-K 稀疏化导致实际效果略弱于此理论值，但多轮迭代下模型仍能有效收敛。综合来看，$\varepsilon = 8.0$ 下精度损失仅为 3.7%是多节点噪声平均、充分训练轮数(100 轮)和较宽松隐私预算三者共同作用的结果。从实用角度看，Jayaraman 和 Evans [24]通过对差分隐私机器学习的系统性实证评估发现，在复杂学习任务中，提供有限精度损失的隐私配置往往仅能提供较弱的有效隐私保护，而提供强隐私保证的配置则导致模型不可用，隐私预算的合理校准仍是一个开放问题。在工业实践中，Apple 在其本地差分隐私部署[25]中针对不同敏感程度的数据类型采用了差异化的隐私预算配置[26]。需要指出的是，本文采用的是中心差分隐私模型，其 $\varepsilon = 8.0$ 的隐私保护强度不能与本地差分隐私模型(如 Apple 部署中每条记录独立加噪)中的 ε 值直接比较。在中心模型下，聚合服务器可观察到加噪后的梯度，隐私风险低于本地模型中数据直接暴露的场景，因此相同 ε 值在中心模

型下提供的实际保护通常更强[21]。本文选取 $\varepsilon = 8.0$ 在当前联邦学习需求预测场景中兼顾了预测精度需求与隐私保护强度，但也需指出，该配置下的隐私保护强度相对宽松，对于涉及高度敏感个人信息的应用场景可能需要进一步降低 ε 值并结合其他隐私增强技术。

4.6. 联邦学习收敛性分析

图 4 记录了联邦训练过程中全局模型在验证集上的 RMSE 变化趋势。EdgeFL-DP 在约 60 轮通信后基本收敛，而 FedAvg-LSTM 需要约 80 轮才能达到相当的收敛水平，且最终精度较低。FedProx-LSTM 的收敛速度介于两者之间。EdgeFL-DP 不仅在最终收敛精度上显著优于 FedAvg-LSTM (RMSE 降低约 15%，该降幅包含双分支架构和 FedAdapt 聚合策略的综合贡献)，还将收敛速度提升了约 25%。其中 FedAdapt 聚合策略通过自适应权重调整有效缓解了 Non-IID 数据的负面影响(消融实验显示其单独贡献约 7.6%的 RMSE 降幅)。这一加速效应源于 FedAdapt 通过分布相似度因子抑制了与全局分布偏差较大的节点的贡献，避免了聚合过程中的梯度冲突。并且对比 SCAFFOLD-DualBranch 和 FedNova-DualBranch 的收敛曲线。SCAFFOLD 得益于方差缩减机制，收敛速度快于 FedNova 和 FedAvg，但最终精度仍不及 EdgeFL-DP，表明梯度校正与自适应聚合权重两种策略的作用层面不同，前者侧重优化方向的校正，后者侧重节点贡献度的差异化调配。

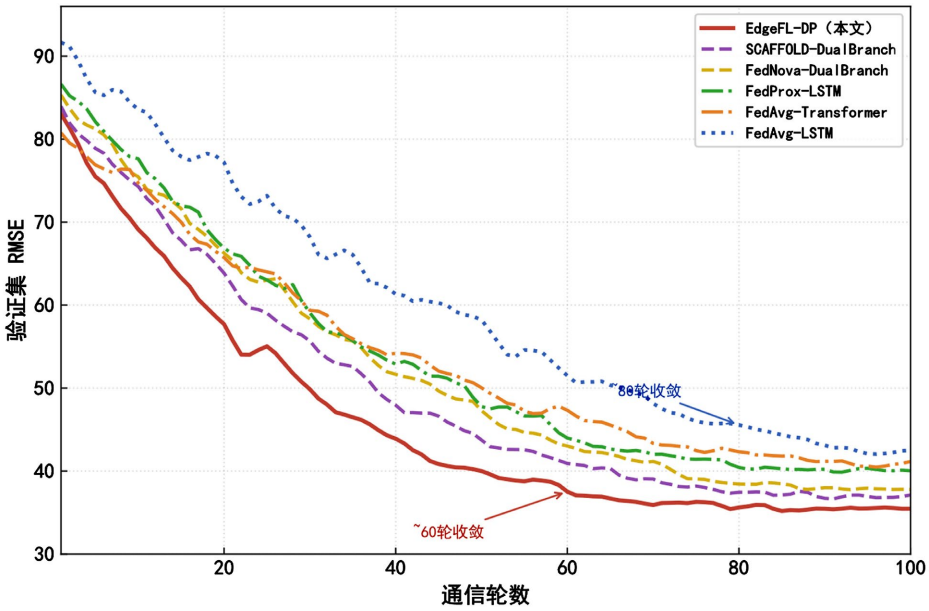

Figure 4. Comparison of federated learning convergence curves
图 4. 联邦学习收敛曲线对比

4.7. 不同节点数量对性能的影响

为评估系统的可扩展性，本文测试了参与联邦训练的边缘节点数量从 4 个增加至 16 个时的性能变化，结果如图 5 所示。随着节点数量的增加，全局模型的预测精度呈先升后趋平的趋势：4 个节点时 RMSE 为 37.82，8 个节点时降至 34.98，12 个节点时为 34.25，16 个节点时为 34.08。这表明更多的参与节点能够提供更丰富的数据多样性，但边际收益在 8 个节点后趋于饱和。通信开销随节点数量近似线性增长(从 185 MB 增加至 705 MB)，但每个节点的计算负载保持不变，体现了框架良好的水平扩展能力。综合考虑精度收益与通信代价，8~12 个节点是推荐的部署规模。

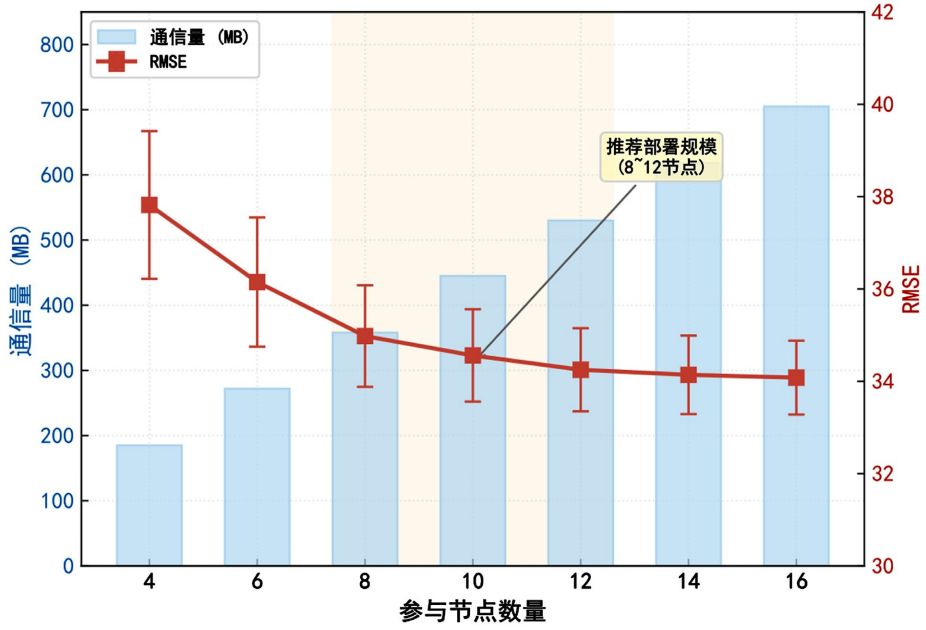

Figure 5. Impact of node count on system performance
图 5. 节点数量对系统性能的影响

5. 讨论

本文提出的 EdgeFL-DP 框架在预测精度、通信效率和隐私保护三个维度上均展现出良好的综合性能。相比集中式单分支基线方法，EdgeFL-DP 在预测精度上取得了显著提升(RMSE 降低 17.8%~23.6%)，该提升是双分支架构(消融实验显示贡献约 6.0%)和 FedAdapt 聚合策略(贡献约 7.6%)等多因素综合作用的结果。在通信效率方面，联邦学习范式将通信内容从原始数据转变为模型梯度，Top-K 稀疏化策略在此基础上将梯度传输量进一步压缩 85.1%，两者共同使数据传输量相比集中式方案减少了 87.4%。在隐私保护方面，差分隐私机制在 $\varepsilon = 8.0$ 的预算下仅引入 3.7%的精度损失，这得益于多节点聚合对独立噪声的平均效应、充分的训练轮数以及相对宽松的隐私预算配置。

然而，本文仍存在以下局限性和未来改进方向：

第一，当前实验基于模拟数据集进行，虽然数据生成过程充分考虑了真实电商场景的关键特征(季节性、促销效应、区域异质性等)，但模拟数据与真实业务数据之间仍存在差距。未来工作将在公开的真实电商数据集(如 Kaggle M5 预测竞赛数据集等)上进行验证，进一步检验框架的泛化能力。

第二，本文的实验设计侧重于评估 EdgeFL-DP 作为完整系统方案相对于现有主流部署方案(集中式单分支训练、标准联邦学习)的综合优势，而非逐一剥离每个技术组件在不同训练范式下的独立贡献。消融实验(表 2)已在联邦框架内部量化了各组件的边际贡献(双分支融合约 6.0%、FedAdapt 约 7.6%)，但由于集中式基线采用单分支架构，表 1 中报告的 17.8%~23.6% RMSE 降幅包含了模型架构差异和联邦聚合策略差异两方面因素。此外，表 1 中通过 SCAFFOLD-DualBranch 和 FedNova-DualBranch 与 EdgeFL-DP 的直接对比，在相同模型架构下验证了 FedAdapt 相对于其他先进 Non-IID 聚合算法的额外优势。严格量化联邦训练范式本身(在相同模型架构下)相对于集中式训练的性能边界，是一个独立的研究问题，值得在后续工作中通过设计集中式双分支对照实验深入探讨。此外，FedAdapt 中的超参数 λ_1、λ_2、λ_3 目前通过验证集网格搜索确定($\lambda_1 = 0.4$、$\lambda_2 = 0.3$、$\lambda_3 = 0.3$)，未来可探索自适应学习这些超参数的方法，以降低人工调参成本并进一步提升算法的通用性。

第三，本文假设各边缘节点均为诚实参与方，未考虑恶意节点或拜占庭攻击的情况。在实际部署中，可以引入基于声誉的节点选择机制或拜占庭容错聚合算法(如 Krum、Trimmed Mean 等)以增强系统的鲁棒性。

第四，差分隐私机制在低隐私预算($\varepsilon \leq 2.0$)下对预测精度造成了较大影响(RMSE 增幅超过 30%)。未来可以探索更高效的隐私保护技术，如安全多方计算或同态加密，或采用自适应隐私预算分配策略，在训练后期逐步降低噪声强度以提升最终精度。

6. 结论

本文针对电商供应链中本地化需求预测面临的数据隐私、通信效率和预测精度挑战，提出了边缘智能驱动的联邦学习需求预测框架 EdgeFL-DP。通过双分支时序预测模型、FedAdapt 加权联邦聚合算法、差分隐私保护机制和梯度 Top-K 稀疏化策略的协同设计，在模拟数据集上实现了较对比方法显著的性能提升。实验结果表明，EdgeFL-DP 的 RMSE 相比集中式单分支 Transformer 基线降低了 17.8%，相比集中式单分支 LSTM 基线降低了 23.6%；消融实验在联邦框架内部量化了各组件的边际贡献，其中双分支架构贡献约 6.0%的 RMSE 降幅，FedAdapt 聚合策略贡献约 7.6%的 RMSE 降幅。在通信效率方面，通过梯度 Top-K 稀疏化将数据传输量减少了 87.4%，通信延迟降低了 65.2%。在 $\varepsilon = 8.0$ 的隐私预算下，借助多节点聚合的噪声平均效应和充分的训练轮数，隐私保护引入的精度损失仅为 3.7%。本研究为边缘环境下的电商供应链智能化管理提供了一种兼顾隐私保护、通信效率和预测精度的系统性技术方案。

参考文献

[1] Statista (2024) Retail E-Commerce Sales Worldwide from 2022 to 2028. https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/

[2] 施巍松, 张星洲, 王一帆, 张庆阳. 边缘计算: 现状与展望[J]. 计算机研究与发展, 2019, 56(1): 69-89.

[3] McMahan, B., Moore, E., Ramage, D., *et al*. (2017) Communication-Efficient Learning of Deep Networks from Decentralized Data. *Proceedings of the* 20*th International Conference on Artificial Intelligence and Statistics* (*AISTATS*), Fort Lauderdale, 20-22 April 2017, 1273-1282.

[4] 李杰, 王玉霞, 赵旭东. 电商企业商品销量的预测方法[J]. 统计与决策, 2018, 34(22): 176-179.

[5] 李建斌, 雷鸣颢, 戴宾, 蔡学媛. 考虑促销因素的医药电商平台需求预测研究[J]. 中国管理科学, 2022, 30(12): 120-130.

[6] 何喜军, 马珊, 武玉英, 蒋国瑞. 小样本下多维指标融合的电商产品销量预测[J]. 计算机工程与应用, 2019, 55(15): 177-184.

[7] 王雪蓉, 万年红. 基于跨境电商可控关联性大数据的出口产品销量动态预测模型[J]. 计算机应用, 2017, 37(4): 1038-1043, 1050.

[8] 程开明, 刘书成, 雷洛, 陈晓颖. 基于网络搜索数据和深度神经网络的社会消费品零售总额预测研究[J]. 运筹与管理, 2024, 33(12): 203-209.

[9] 梁宏涛, 刘硕, 杜军威, 胡强, 于旭. 深度学习应用于时序预测研究综述[J]. 计算机科学与探索, 2023, 17(6): 1285-1300.

[10] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

[11] Vaswani, A., Shazeer, N., Parmar, N., *et al*. (2017) Attention Is All You Need. *Advances in Neural Information Processing Systems*, **30**, 5998-6008.

[12] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., *et al*. (2021) Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, **35**, 11106-11115. https://doi.org/10.1609/aaai.v35i12.17325

[13] Salinas, D., Flunkert, V., Gasthaus, J. and Januschowski, T. (2020) DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks. *International Journal of Forecasting*, **36**, 1181-1191. https://doi.org/10.1016/j.ijforecast.2019.07.001

[14] Rangapuram, S.S., Seeger, M.W., Gasthaus, J., *et al*. (2018) Deep State Space Models for Time Series Forecasting. *Proceedings of the* 32*nd International Conference on Neural Information Processing Systems*, Montréal, 3-8 December 2018, 7796-7805.

[15] 周传鑫, 孙奕, 汪德刚, 葛桦玮. 联邦学习研究综述[J]. 网络与信息安全学报, 2021, 7(5): 77-92.

[16] Li, T., Sahu, A.K., Zaheer, M., *et al*. (2020) Federated Optimization in Heterogeneous Networks. *Proceedings of Machine Learning and Systems*, **2**, 429-450.

[17] Li, Q., He, B. and Song, D. (2021) Model-Contrastive Federated Learning. 2021 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, 20-25 June 2021, 10713-10722. https://doi.org/10.1109/cvpr46437.2021.01057

[18] Li, X., Jiang, M., Zhang, X., *et al*. (2021) FedBN: Federated Learning on Non-IID Features via Local Batch Normalization. *Proceedings of the* 9*th International Conference on Learning Representations*, Virtual Event, 3-7 May 2021.

[19] Karimireddy, S.P., Kale, S., Mohri, M., *et al*. (2020) SCAFFOLD: Stochastic Controlled Averaging for Federated Learning. *Proceedings of the* 37*th International Conference on Machine Learning*, Virtual Event, 13-18 July 2020, 5132-5143.

[20] Wang, J., Liu, Q., Liang, H., *et al*. (2020) Tackling the Objective Inconsistency Problem in Heterogeneous Federated Optimization. *Advances in Neural Information Processing Systems*, **33**, 7611-7623.

[21] Dwork, C. and Roth, A. (2014) The Algorithmic Foundations of Differential Privacy. *Foundations and Trends® in Theoretical Computer Science*, **9**, 211-487. https://doi.org/10.1561/0400000042

[22] Abadi, M., Chu, A., Goodfellow, I., McMahan, H.B., Mironov, I., Talwar, K., *et al*. (2016) Deep Learning with Differential Privacy. *Proceedings of the* 2016 *ACM SIGSAC Conference on Computer and Communications Security*, Vienna, 24-28 October 2016, 308-318. https://doi.org/10.1145/2976749.2978318

[23] 李敏, 肖迪, 陈律君. 兼顾通信效率与效用的自适应高斯差分隐私个性化联邦学习[J]. 计算机学报, 2024, 47(4): 924-946.

[24] Jayaraman, B. and Evans, D. (2019) Evaluating Differentially Private Machine Learning in Practice. *Proceedings of the* 28*th USENIX Conference on Security Symposium*, Santa Clara, 14-16 August 2019, 1895-1912.

[25] Apple Differential Privacy Team (2017) Learning with Privacy at Scale. https://machinelearning.apple.com/research/learning-with-privacy-at-scale

[26] Apple (2017) Differential Privacy Overview. https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf