

DOI:10.11784/tdxbz202501001

基于联邦学习与知识蒸馏的轻量化负荷分解方法

王守相¹, 曹 智¹, 赵倩宇¹, 冯喜春², 容春艳²

(1. 天津大学智能电网教育部重点实验室, 天津 300072;

2. 国网河北省电力有限公司经济技术研究院, 石家庄 050021)

摘要: 针对深度学习模型在非侵入式负荷分解中面临的数据隐私保护和边缘部署两个问题, 提出了一种基于联邦学习与知识蒸馏的轻量化负荷分解框架与方法。首先, 设计了一种结合卷积神经网络(CNN)和 Transformer(CNN-Transformer)的混合架构, 通过 CNN 模块高效提取负荷序列的局部时序特征, 利用改进的 Transformer 结构增强对长期时序依赖关系的建模能力, 提高了模型的整体辨识性能; 其次, 提出基于知识蒸馏的模型轻量化策略, 通过设计知识迁移机制, 将大参量教师模型的决策能力有效压缩至轻量级学生模型, 实现模型的高效轻量化; 最后, 构建了基于联邦学习-知识蒸馏的云边协同训练架构, 采用联邦平均算法实现模型参数的高效聚合, 使边缘节点在不共享原始数据的情况下参与模型训练, 同时引入轻量化模型作为全局模型显著降低了通信开销。实验结果表明: 所提模型在 REDD 和 UK-DALE 数据集上的辨识性能优于现有方法; 轻量化策略在将模型参数降低 90% 的同时保持了较好的精度; 所提框架较传统联邦学习降低了约 85% 的通信量, 为非侵入式负荷分解在边缘计算场景中的部署提供了有效方案。

关键词: 非侵入式负荷分解; 联邦学习; 知识蒸馏; 轻量化; 隐私保护

中图分类号: TM71

文献标志码: A

文章编号: 0493-2137(2026)01-0052-13

A Lightweight Load Monitoring Method Based on Federated Learning and Knowledge Distillation

Wang Shouxiang¹, Cao Zhi¹, Zhao Qianyu¹, Feng Xichun², Rong Chunyan²

(1. Key Laboratory of Smart Grid of Ministry of Education, Tianjin University, Tianjin 300072, China;

2. Economic and Technological Research Institute, State Grid Hebei Electric Power Co., Ltd.,
Shijiazhuang 050021, China)

Abstract: To address the issues of data privacy protection and edge deployment in deep learning models for non-intrusive load monitoring (NILM), this paper proposes a lightweight load decomposition framework and method based on federated learning and knowledge distillation. First, a hybrid convolutional neural network and Transformer (CNN-Transformer) architecture was designed. The CNN module efficiently extracted local temporal features from load sequences, while the improved Transformer structure enhanced the modeling capability of long-term temporal dependencies, thus improving the overall identification performance. Second, a model lightweight strategy based on knowledge distillation was proposed. Through a designed knowledge transfer mechanism, the decision-making capability of the large-parameter teacher model was effectively compressed into a lightweight student model, achieving efficient model compression. Finally, a cloud-edge collaborative training architecture based on federated learning and knowledge distillation was constructed. The federated averaging algorithm enabled efficient aggregation of model parameters, allowing edge nodes to participate in model training without sharing raw data. The introduction of the lightweight model as the global model significantly reduced communication overhead. Experimental results

收稿日期: 2025-01-01; 修回日期: 2025-03-07.

作者简介: 王守相 (1973—), 男, 博士, 教授, sxwang@tju.edu.cn.

通信作者: 赵倩宇, zhaoqianyu@tju.edu.cn.

基金项目: 国家电网公司总部科技资助项目 (SGHEJY00PSJS2310023).

Supported by the Science and Technology Project of State Grid Corporation of China (No. SGHEJY00PSJS2310023).

show that the proposed model outperforms existing methods on REDD and UK-DALE datasets. The lightweight strategy maintains good accuracy while reducing model parameters by 90%. Compared to traditional federated learning, the proposed framework reduces communication volume by approximately 85%, providing an effective solution for NILM deployment in edge computing scenarios.

Keywords: non-invasive load monitoring (NILM); federated learning; knowledge distillation; lightweight; privacy protection

近年来,随着电力系统向智能化方向的深入发展,配电网对数据采集、分析与态势感知的需求日益增长^[1]。作为新型电力基础设施的重要组成部分,智能电表与边缘计算融合终端的广泛部署不仅提供了丰富的用电数据,也为负荷分析在边缘侧的本地化部署创造了有利条件^[2]。然而,传统的负荷分解技术需要在每个用电设备处安装独立的测量装置,这种方式不仅增加了系统部署和维护成本,也限制了其在实际场景中的推广应用^[3-4]。

非侵入式负荷分解(non-intrusive load monitoring, NILM)通过对用户总电力负荷的分析来辨识单个用电设备的运行状态及其能耗特征^[5-6]。NILM 技术将总负荷数据分解为各设备用电信息,实现了用电行为的精细化识别,使负荷分析能够深入到设备级别,为负荷预测、异常检测和需求响应策略制定提供了更精准的数据支撑。作为电力系统负荷分析的关键技术, NILM 能够为配电网优化规划和需求侧管理提供坚实的数据基础,提升电力系统运行效率和用能管理水平。与传统负荷分解技术相比, NILM 无需额外安装测量设备,具有显著的经济性与实用性优势^[7]。传统 NILM 方法主要依赖人工设计的特征提取与模式匹配算法,在处理复杂负荷模式时存在辨识瓶颈^[8-9]。近年来,深度学习技术在多个领域取得重大突破,其强大的特征提取与模式辨识能力为 NILM 技术的发展开创了新的路径^[10-11]。基于深度学习的 NILM 方法展现出优异的负荷分解性能,逐渐成为该领域的研究重点。目前,针对深度学习在 NILM 领域的应用已开展了大量研究,文献[12]针对神经网络 NILM 方法在负荷分解精度上的局限性,提出了基于生成对抗网络(generative adversarial network, GAN)的改进方法,但性能提升不显著。文献[13]通过格拉姆角场将低频功率数据转换为二维图像,有效提升了模型的特征提取能力和泛化性能。文献[14]将 U-Net 架构与实例归一化技术相结合,提出基于条件 GAN 的序列到子序列方法,实现了对多尺度特征的有效捕获。文献[15]提出了基于多通道融合机制的方法,通过融合不同特征提高了模型的整体辨识性能。文献[16]将 $V-I$ 轨迹特征与功率特征相结合,提升了负荷分解精度。

上述研究主要聚焦于模型结构优化和特征融合策略以提升负荷分解精度,但对模型轻量化部署和数据隐私保护等实际问题的研究相对不足。在实际应用中,基于深度学习的 NILM 技术面临两个关键挑战。首先,高性能深度学习模型的训练依赖大规模数据集。随着社会对数据隐私保护重视程度的提高,国家相继出台了数据保护法规,这使得传统的集中式数据采集与模型训练方式难以为继^[17],亟需探索符合隐私保护要求的分布式训练方案^[18]。其次,为获得较高的负荷分解精度,深度学习模型往往采用复杂的网络结构与庞大的参数规模。这种计算密集型模型与边缘侧设备在计算能力、存储空间和通信带宽等资源上的限制之间存在本质矛盾,需要研究在保持模型性能的前提下实现轻量化部署的有效方法^[19]。

联邦学习作为一种分布式机器学习范式^[20],通过本地训练与参数聚合的方式实现了数据隐私保护下的模型协同优化,为解决 NILM 技术的数据隐私问题提供了有效途径。文献[21]设计了一种多层次隐私保护联邦学习框架,有效改善了 NILM 在联邦学习环境中隐私保护与训练效果难以平衡的问题。文献[22]将垂直联邦学习与 NILM 相结合,对家用电器数据进行清洗与优化,以提高模型训练的数据质量,同时保障用户数据隐私安全。然而,现有的 NILM 联邦学习方法主要关注于隐私保护机制的设计,较少考虑模型本身在计算资源受限的边缘设备上的部署效率,难以平衡隐私保护、模型性能与计算资源限制的多重约束。知识蒸馏技术提供了一种将复杂模型的决策能力迁移至轻量化模型的有效途径^[23],为解决 NILM 技术中的边缘部署问题提供了新的思路。

针对上述问题,本文提出了一种基于联邦学习与知识蒸馏的轻量化负荷分解框架,该框架通过联邦学习实现隐私保护下的分布式训练,并利用知识蒸馏技术将模型轻量化,同时设计了一种改进的 Transformer 模型以进一步提升轻量化模型的辨识性能。首先,设计了一种基于卷积神经网络(convolutional neural network, CNN)和 Transformer (CNN-Transformer) 混合架构的改进模型,通过 CNN 模块提取负荷序列的局部特征模式,利用改进的 Transformer 结构增强

对长期时序依赖关系的建模能力,提高模型的整体辨识性能.其次,构建了基于联邦学习与知识蒸馏的轻量化框架,利用知识蒸馏技术将复杂模型的决策能力迁移至轻量化模型,通过联邦学习机制实现数据隐私保护下的分布式训练.最后,基于 REDD 和 UK-DALE 两个公开数据集对所提模型与框架进行了全面验证,从模型辨识性能、参数规模、计算效率和通信开销等多个维度进行了对比分析.

1 改进 Transformer 模型及其原理

1.1 Transformer 模型

Transformer 广泛应用于各类深度学习任务中,其架构基于编码器-解码器(encoder-decoder)进行构建.编码器部分通常由多个相同的编码器层堆叠而成,每一个编码器层包含两个核心子层.第 1 个子层是注意力机制,第 2 个子层是位置前馈神经网络.两个子层都采用了残差连接与层归一化.解码器的架构与编码器类似,其主要的区别在于解码器中有两个多头注意力机制子层.

1.1.1 注意力机制

注意力机制的核心是通过计算查询矩阵 Q 与键值矩阵 K 的相似度来确定注意力的分布,然后根据计算得来的分布对值矩阵 V 进行加权求和.计算过程主要包含 3 步:首先计算查询矩阵 Q 和键矩阵 K 的点积 $Q \cdot K^T$;然后使用缩放因子进行缩放并通过 Softmax 函数将所得数值转化为概率分布;最后利用这个概率分布对值矩阵 V 进行加权求和,得到最终的注意力输出.注意力机制的标准形式可以表示为

$$A(Q, K, V) = \text{Softmax}(QK^T / \sqrt{d})V \quad (1)$$

式中: Q 为查询矩阵; K 为键值矩阵; V 为值矩阵; d 为 3 个矩阵的维度

1.1.2 位置前馈神经网络

除注意力子层之外,编码器、解码器层中还包含一个相同的位置前馈神经网络子层.在传统 Transformer 架构中,该子层的处理流程如下:首先利用一个线性层将输入特征投影到更高维的特征空间中,随后使用 ReLU 激活函数引入非线性变换能力,最后通过另一个线性层将特征重新映射回原始维度.这个过程可以表示为

$$F(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (2)$$

式中: x 代表输入向量; W_1 、 W_2 代表两个可学习的权重矩阵; b_1 、 b_2 为权重矩阵对应的偏置向量.

1.2 改进 Transformer 模型

本文设计了一种面向非侵入式负荷分解的混合 Transformer 模型,其架构如图 1 所示.原始的 Transformer 结构通过其多头注意力机制能够有效地分析与处理长序列时序数据,有效捕获数据中的时间依赖特征.然而使用 Transformer 模型直接处理原始序列数据,存在着模型对局部时序特征提取能力有限的问题.针对这一问题,本文在 Transformer 架构的基础上引入了 CNN 结构, CNN 利用其特有的卷积运算方式,能够充分提取时序数据中的局部特征模式,对于电器负载特征的识别具有显著优势.通过融合这两个架构,利用 CNN 对原始数据进行提取,深化模型对于时序特征的提取能力,丰富了模型所能获取与处理的时序特征信息,随后采用 Transformer 架构对提取的特征进行深层次的时序建模,实现了对多尺度用电特征的全面捕获,提升了模型在负荷分解任务中的性能表现.

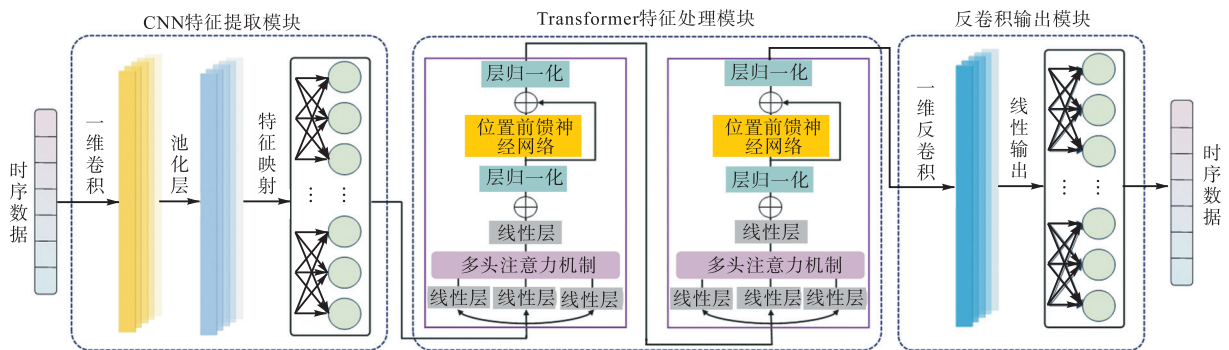


图 1 改进 Transformer 架构
Fig.1 Architecture of improved Transformer

1.2.1 CNN 模块

CNN 模块主要由特征卷积层、池化层和特征投影层 3 个部分组成.该模块的主要目的是对输入的

负荷功率序列进行特征提取和维度转换,为后续的 Transformer 模块提供更具表征能力的特征输入.

首先,特征卷积层采用一维卷积运算,通过局部

感受野机制提取负荷功率序列的时序特征为

$$\mathbf{F}_{\text{conv}}(\mathbf{x}_1) = \mathbf{W}_c \times \mathbf{x}_1 + \mathbf{b}_c \quad (3)$$

式中： $\mathbf{x}_1$ 代表输入的负荷功率序列； $\mathbf{W}_c$ 为卷积核参数矩阵； $\mathbf{b}_c$ 为卷积核矩阵的偏置项。

该层将输入序列映射到高维特征空间，有效捕获了负荷功率的局部变化特征。

然后，为了降低计算复杂度并提取更紧凑的特征表示，将卷积层的输出送入 LP 池化层进行二范数 (L2-norm) 池化操作，即

$$\mathbf{F}_{\text{pool}}(\mathbf{F}_{\text{conv}}) = \sqrt{\sum_{i=1}^s \mathbf{F}_{\text{conv},i}^2} \quad (4)$$

式中： $\mathbf{F}_{\text{conv}}$ 为特征卷积层的输出； s 为池化核大小； $\mathbf{F}_{\text{conv},i}$ 表示池化窗口内第 i 个位置的特征值。

该操作在保留信号能量特征的同时，将序列长度从 480 压缩至 240，显著降低了后续处理的计算量。

最后，特征投影层通过线性变换将池化后的特征重排为适合 Transformer 处理的格式，即

$$\mathbf{F}_{\text{proj}}(\mathbf{F}_{\text{pool}}) = \mathbf{W}_p \mathbf{F}_{\text{pool}} + \mathbf{b}_p \quad (5)$$

式中： $\mathbf{F}_{\text{pool}}$ 为池化层输出； $\mathbf{W}_p$ 为投影矩阵； $\mathbf{b}_p$ 为投影矩阵对应的偏置项。

CNN 模块不仅提取了负荷功率序列的局部模式特征，还为后续的 Transformer 模块提供了结构化的特征表示，从而增强了模型对负荷分解任务中时序特征的建模能力。

1.2.2 Transformer 模块

相较于传统 Transformer 的编码器-解码器架构，本模型仅保留编码器部分，通过精简网络结构来降低计算复杂度的同时保持模型的特征提取能力。编码器由两个相同的 Transformer 块堆叠而成，每个块包含多头注意力层和前馈网络两个子模块，使得模型保持了对时序特征的建模能力，同时避免了解码器带来的额外计算开销。

在每个 Transformer 块中，多头注意力层和前馈网络后都使用了残差连接和层归一化操作。对于子模块输出 $S(\mathbf{x})$ ，其处理过程为

$$\mathbf{x}_2 = L_N(\mathbf{x} + S(\mathbf{x})) \quad (6)$$

式中： $\mathbf{x}_2$ 代表残差连接与层归一化处理后的输出向量； $S(\mathbf{x})$ 表示子模块（多头注意力层或前馈网络）的输出。

层归一化操作定义为

$$L_N(\mathbf{x}) = \gamma \odot \frac{\mathbf{x} - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta_1 \quad (7)$$

式中： μ 和 σ 分别为特征均值和标准差； γ 和 ϵ 为可学习的缩放和平移参数； β_1 为防止除零的小常数，取

值为 1×10^{-6} ； $\odot$ 表示逐元素相乘。层归一化有效缓解了深度网络的梯度问题，改善了模型训练的稳定性。

在前馈网络部分，本模型采用 GELU 激活函数替代传统的 ReLU，即

$$G(\mathbf{x}) = 0.5 \mathbf{x} \left\{ 1 + \tanh \left[\sqrt{2/\pi} (\mathbf{x} + 0.044715 \mathbf{x}^3) \right] \right\} \quad (8)$$

式中 $\tanh$ 为双曲正切函数。 G 函数通过引入非线性变换和平滑过渡，能够提升模型的非线性拟合能力。相比 ReLU 的强截断特性， G 在负值区域表现出更平滑的梯度变化，有助于模型捕捉负荷时序数据中的细微模式。前馈网络的完整计算过程为

$$\mathbf{F}(\mathbf{x}) = \mathbf{W}_2 (G(\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1)) + \mathbf{b}_2 \quad (9)$$

式中 $\mathbf{W}_1$ 和 $\mathbf{W}_2$ 均为权重矩阵， $\mathbf{W}_1 \in \mathbf{R}^{256 \times 1024}$ ， $\mathbf{W}_2 \in \mathbf{R}^{1024 \times 256}$ 。

本模型采用多头注意力机制对 CNN 提取的高层时序特征进行建模。对于单个注意力头，其计算过程与前文所述注意力机制计算方法相同。为了充分利用 CNN 提取的特征信息并增强特征表示的多样性，模型采用的双头注意力机制 ($h=2$) 为

$$\mathbf{M}(\mathbf{x}) = C(h_1, h_2) \mathbf{W}^0 \quad (10)$$

$$h_i = A(\mathbf{x} \mathbf{W}_i^Q, \mathbf{x} \mathbf{W}_i^K, \mathbf{x} \mathbf{W}_i^V) \quad (11)$$

式中： $\mathbf{W}_i^Q$ 、 $\mathbf{W}_i^K$ 、 $\mathbf{W}_i^V \in \mathbf{R}^{256 \times 128}$ 为第 i 个注意力头的参数矩阵； $\mathbf{W}^0 \in \mathbf{R}^{256 \times 256}$ 为输出投影矩阵。通过将多头注意力机制应用于 CNN 提取的特征表示，模型既保持了对局部时序模式的感知能力，又能够从多个特征子空间建立长程时序依赖关系，从而提升了对复杂负荷模式的建模能力。

2 联邦学习-知识蒸馏轻量化架构

2.1 联邦学习概述

联邦学习是一种分布式机器学习框架，其核心思想是通过在多个本地数据源上进行分布式模型训练，同时仅交换模型参数而非原始数据，从而实现数据隐私保护与模型性能之间的平衡。该框架旨在解决人工智能模型训练中的数据孤岛问题。传统的集中式训练方法要求将分散的用户数据集中到单一计算节点，这不仅存在数据传输过程中的安全风险，还面临用户因隐私顾虑而不愿共享数据的挑战。联邦学习通过允许模型在本地训练并仅共享模型参数的方式，既保护了用户隐私，又能充分利用分散数据来提升模型性能。

联邦学习旨在通过最小化所有参与方的平均损失函数来优化全局模型，即

$$\min_{w \in \mathbf{R}} f(w) = \sum_{k=1}^K \frac{n_k}{n} F_k(w) \quad (12)$$

式中: $F_k(w)$ 为参与方 k 的局部损失函数; n_k 为参与方 k 边缘节点的本地数据量; n 为所有参与训练节点的总数据量; K 为参与训练的边缘节点总数.

联邦学习流程包括: 选择全局模型并下发至各参与方, 参与方利用本地私有数据集进行训练直至收敛或满足特定条件; 随后各参与方将训练后的模型参数上传至中央服务器, 中央服务器采用联邦聚合算法进行模型更新. 在整个训练过程中, 联邦学习仅对模型参数进行通信与交互, 各参与方的私有数据始终保持在本地, 既没有提供给模型训练商, 也没有产生任何通信, 从而有效保护了数据隐私.

在实际应用中, 不同参与方的数据在样本量、特征维度和分布特性等方面存在显著差异. 为适应这种数据异构性, 发展出了不同结构的联邦学习框架. 横向联邦学习适用于特征维度一致但样本标识不同的场景, 通过聚合不同来源的样本信息在保护数据隐私的前提下提升模型的泛化能力. 本文联邦学习架构基于非侵入式负荷分解场景构建, 考虑到不同本地数据集具有相似的特征维度, 联邦学习框架用于整合分散的用电数据, 在确保用户隐私不泄露的前提下扩充样本多样性并提升模型性能, 因此选择横向联邦学习模式.

2.2 知识蒸馏概述

随着深度学习模型规模不断扩大, 在边缘计算设备等资源受限场景下的部署问题日益突出. 知识蒸馏技术是一种高效模型轻量化方法, 其核心思想是通过设计知识迁移机制, 将大规模教师模型中的决策能力和特征表示有效压缩到轻量级学生模型中, 其架构如图 2 所示, 图中 T 为温度参数, t 为实际训练过程中选择的超参数, 当 $T=1$ 时模型输出原始预测结果.

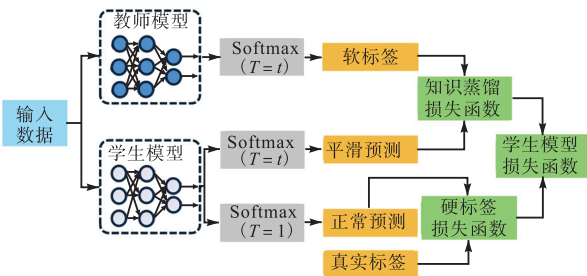


图 2 知识蒸馏架构

Fig.2 Architecture of knowledge distillation

知识蒸馏框架主要包含 3 个关键组件: 教师模型、学生模型和知识迁移模块. 教师模型采用复杂网络结构, 具有强大的特征提取和表示能力; 学生模型

采用轻量化架构, 通过知识迁移模块获取教师模型的知识指导. 通过这种知识迁移机制, 实现了模型的高效轻量化部署, 在显著降低计算复杂度和存储开销的同时, 充分利用教师模型的知识指导来增强学生模型的泛化能力和鲁棒性, 从而在资源受限场景下仍能保持较好的性能表现.

2.3 联邦学习-知识蒸馏框架

2.3.1 框架架构

本文所提联邦学习-知识蒸馏架构如图 3 所示, 该架构融合了联邦学习与知识蒸馏两项技术. 在联邦学习分布式隐私保护模型训练的基础上, 引入知识蒸馏的轻量化策略, 将传统联邦学习过程中的全局模型迭代替换为轻量化模型迭代, 有效降低了云边协同过程中的通信开销, 并且实现了模型在边缘侧的轻量化部署. 整个架构分为云端两端, 其中云端的中央服务器部署了知识蒸馏与联邦平均两个核心算法. 云端服务器基于内置的预训练大模型, 采用知识蒸馏技术对学生模型进行训练, 并将学生模型通过联邦学习模式下发到边缘端进行迭代. 边缘节点可以根据自身实际情况灵活选择是否加入联邦学习架构, 以及是否参与协助全局模型的训练过程.

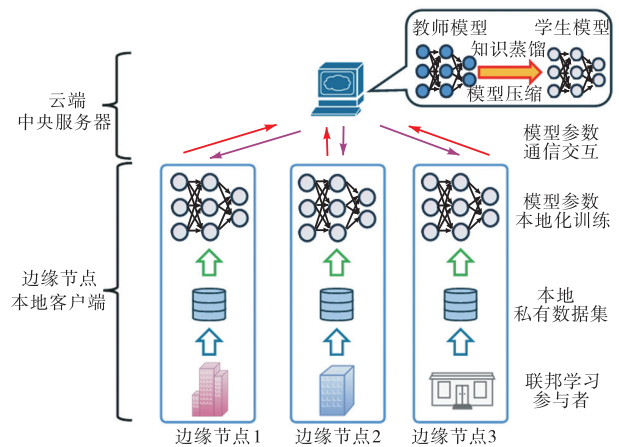


图 3 联邦学习-知识蒸馏联合架构

Fig.3 Joint architecture of federated learning-knowledge distillation

本文所提架构的工作流程主要包含如下阶段.

首先, 在联邦学习的云端中心节点部署预训练好的教师模型, 同时收集非侵入式负荷分解的公共数据集, 并对数据集进行划分与处理.

其次, 云端中心基于非侵入式负荷分解任务需求, 构建轻量化学生模型并初始化学生模型参数

$$W = w^{(0)} \quad (13)$$

针对知识蒸馏任务, 构建学生模型的知识蒸馏损失函数

$$L_s = \alpha L_{CE}(y_s, y) + (1 - \alpha) L_{KD}(z_t, z_s) \quad (14)$$

式中：\$L_s\$ 代表学生模型的总损失函数；\$L_{CE}(y_s, y)\$ 为学生模型原本的损失函数，\$y_s\$ 为学生模型输出的硬标签，\$y\$ 为真实标签；\$L_{KD}(z_t, z_s)\$ 为知识蒸馏损失函数，\$z_t\$ 为教师模型经过温度系数处理后输出的软标签；\$z_s\$ 为学生模型经过温度系数处理后输出的软标签；\$\alpha\$ 为权重系数，用于平衡原任务学习与知识蒸馏。

完成知识蒸馏后，将优化的学生模型作为初始全局模型分发至各边缘节点。边缘节点接收模型后，可以选择直接部署该模型，或基于本地数据集进行进一步训练。

学生模型处于边缘侧计算时不再进行知识蒸馏过程，此时训练过程仅关注模型的基础损失函数，其损失函数简化为

$$L_s = L_{CE}(y_s, y) \quad (15)$$

参与联邦学习的边缘节点完成本地训练后，通过梯度下降法更新模型参数。在第 \$i\$ 次本地训练迭代中，模型参数按照更新方式为

$$w_{m+1}^i = w_m^i - \beta_2 \nabla L_{CE}(w_m^i) \quad (16)$$

式中：\$w_m^i\$ 表示第 \$i\$ 次迭代的模型参数；\$\beta_2\$ 为学习率；

\$\nabla L_{CE}(w_m^i)\$ 为损失函数关于当前参数的梯度；\$m\$ 为当前联邦学习迭代的轮次。

最后，云端中心节点汇集所有参与训练的边缘节点上传的模型参数，采用联邦平均 (federated average, FedAvg) 算法进行全局参数聚合，即

$$w_{m+1} = \sum_{k=1}^K \frac{n_k}{n} w_m^k \quad (17)$$

通过联邦聚合算法，生成新一轮的全局模型为

$$W = w^{(m)} \quad (18)$$

该全局模型将被再次分发至参与联邦学习的边缘节点进行新一轮训练，直至达到预设的迭代轮次。

2.3.2 教师模型选择

在知识蒸馏的过程中，教师模型往往选择参数量较大、性能优越的深度神经网络。因此，本文采用第 1 节所提的改进 Transformer 模型作为教师模型。该模型具有较强的特征提取和时序建模能力，能够为知识蒸馏提供充分的特征知识和决策信息，有效指导学生模型的训练过程。所提模型的参数如表 1 所示。经过计算，该模型的模型参数约为 200 万，实际模型文件大小约为 15 MB。

表 1 改进 Transformer 模型参数

Tab.1 Model parameters of improved Transformer

层名称	输出尺寸	结构设置
特征卷积	480×256	一维卷积层(输入通道 = 1, 输出通道 = 256, 卷积核 = 5, 步长 = 1, 填充因子 = 2)
池化层	240×256	一维池化层(L2 范数, 池化核 = 2, 补偿 = 2)
特征投影	240×256	线性层(输入通道 = 256, 输出通道 = 256)
Transformer 块 1	240×256	多头注意力机制(注意力头 = 2, 注意力维度 = 128)+前馈神经网络(1 024)
Transformer 块 2	240×256	多头注意力机制(注意力头 = 2, 注意力维度 = 128)+前馈神经网络(1 024)
特征反卷积	480×256	一维反卷积层(输入通道 = 256, 输出通道 = 256, 卷积核 = 4, 步长 = 2, 填充因子 = 1)
输出映射	480×1	线性层(输入通道 = 256, 中间通道 = 128, 输出通道 = 1)

同时，为了使所提模型能够适应知识蒸馏任务，更好地融入所提架构当中，需要对教师模型的输出进行软化。在知识蒸馏中，输出软化是一种通过调整模型输出分布来增强知识迁移效果的技术。输出软化能够将教师模型中蕴含的细粒度知识以更温和的形式传递给学生模型，从而实现更有效的知识迁移。

知识蒸馏的输出软化通常采用引入温度参数的 Softmax 函数

$$q_i = \frac{\exp(z_i / T)}{\sum_j \exp(z_j / T)} \quad (19)$$

式中：\$q_i\$ 表示教师模型经过软化后得到的概率分布；\$z_i\$ 代表教师模型最后一层全连接层的输出值。

2.3.3 学生模型选择

为了实现模型的轻量化，学生模型的选择需要考虑模型的简单性和计算效率。本文基于长短期记忆

网络(long short-term memory, LSTM)构建了轻量化学生模型。

所构建的学生模型，首先使用一维卷积层对输入数据进行特征提取

$$F_{conv} = C_{1D}(xW_{conv}) \quad (20)$$

式中：\$F_{conv}\$ 代表卷积层输出特征；\$C_{1D}(x, W_{conv})\$ 为一维卷积；\$W_{conv}\$ 为卷积核参数矩阵。

其次，使用双层 LSTM 来处理序列数据。在第 \$u\$ 个时间步，每层 LSTM 的运算过程为

$$i_u = \sigma(W_{ii}x_u + W_{hi}h_{u-1} + b_i) \quad (21)$$

$$f_u = \sigma(W_{if}x_u + W_{hf}h_{u-1} + b_f) \quad (22)$$

$$g_u = \tanh(W_{ig}x_u + W_{hg}h_{u-1} + b_g) \quad (23)$$

$$o_u = \sigma(W_{io}x_u + W_{ho}h_{u-1} + b_o) \quad (24)$$

$$c_u = f_u \odot c_{u-1} + i_u \odot g_u \quad (25)$$

$$h_u = o_u \odot \tanh c_u \quad (26)$$

式中: i_u 、 f_u 、 o_u 分别为输入门、遗忘门、输出门 u 时刻的状态; g_u 为 u 时刻候选记忆单元; h_u 为 u 时刻隐藏状态; c_u 为 u 时刻单元状态; W_{ii} 、 W_{if} 、 W_{ig} 、 W_{io} 为输入权重矩阵, W_{hi} 、 W_{hf} 、 W_{hg} 、 W_{ho} 为隐藏状态权重矩阵; b_i 、 b_f 、 b_g 、 b_o 为偏置项; σ 为 sigmoid 激活函数。

最后, 模型采用一个线性层输出模型结果, 即

$$\hat{y} = W_{out} h_u + b_{out} \quad (27)$$

表 2 LSTM 模型参数

Tab.2 LSTM model parameters

层名称	输出尺寸	结构设置
特征卷积	480 × 32	一维卷积层(输入通道 = 1, 输出通道 = 32, 卷积核 = 5, 步长 = 1, 填充因子 = 2)
LSTM 层 1	480 × 128	输入通道 = 32, 输出通道 = 128
LSTM 层 2	480 × 128	输入通道 = 128, 输出通道 = 128
输出层	480 × 1	线性层(输入通道 = 128, 输出通道 = 1)

同时, 为使学生模型能够适应知识蒸馏任务, 接受来自教师模型的知识迁移, 同样需要对学生模型的输出进行软化, 并构建知识蒸馏损失函数。学生模型输出软化过程与上文教师模型输出软化过程类似, 此处不再赘述。通常知识蒸馏损失函数采用 KL 散度(Kull-back-Leibler divergence)来度量教师模型和学生模型输出分布之间的差异, 其公式为

$$L_{KD} = KL(P^T \| Q^T) = T^2 \sum_i P^T(i) \lg \left(\frac{P^T(i)}{Q^T(i)} \right) \quad (28)$$

式中: KL 代表 KL 散度; $P^T(i)$ 和 $Q^T(i)$ 分别表示教师模型和学生模型经过温度软化后的输出分布。

3 算例分析

3.1 实验环境及其数据集选择

本文所提框架的实验在配置有 AMD Ryzen 3930X 处理器和 NVIDIA RTX 3090 显卡的 Windows 10 深度学习服务器上进行。实验代码采用 Python 3.9 编写, 深度学习模型与框架使用 PyTorch 2.5.1 构建。

为验证所提模型及联邦学习-知识蒸馏框架的有效性, 本文选取了 REDD 和 UK-DALE 两个公开数据集进行实验。其中, 两个数据集均用于验证模型改进的有效性, 而 UK-DALE 数据集还额外用于验证所提框架的性能。

1) REDD 数据集

REDD 数据集由麻省理工学院(MIT)采集, 包含 6 户家庭的电器功率和总用电功率数据。实验选取冰箱、洗衣机、洗碗机数据验证模型, 将 2~6 号房屋数据作为训练集(其中 10% 为验证集), 1 号房屋数据作为测试集。

式中: $\hat{y}$ 为预测输出; W_{out} 为输出层权重矩阵; b_{out} 为输出层偏置项

表 2 展示了模型的具体参数设置。通过计算, 该模型的总参数量约为 20 万, 实际模型文件大小约为 2 MB。与本文所用教师模型进行对比, 轻量化 LSTM 模型的参数量仅为教师模型的 10%, 具有显著的轻量化优势。轻量化模型不仅降低了存储开销, 也使得模型更容易部署在资源受限的边缘设备上。

2) UK-DALE 数据集

UK-DALE 数据集包含 5 户英国家庭的电器功率和总功率数据。本研究开展两类实验: 一是验证模型有效性, 选取水壶、冰箱、洗衣机和洗碗机数据, 将 1、3、4、5 号房屋数据作为训练集(其中 10% 为验证集), 2 号房屋数据作为测试集; 二是验证联邦学习-知识蒸馏框架, 由于数据非完整性(3 号房屋仅有水壶数据, 4 号房屋仅有水壶和洗衣机数据), 选择水壶和洗衣机进行验证, 将 3、4 号房屋数据作为云端公共数据集, 1 号和 5 号房屋数据分别配置给两个边缘节点作为私有数据集。

3.2 数据预处理

1) 重采样处理

REDD 与 UK-DALE 两个数据集的原始采样频率存在差异: REDD 数据集中总表以 1 s/点、分表以 3 s/点的频率采样, 而 UK-DALE 数据集中总表和分表均以 6 s/点的频率采样。为统一数据格式并保证实验的一致性, 本文将两个数据集统一重采样至 6 s/点的采样频率。

2) 数据清洗与过滤

考虑到实际数据采集过程中可能存在的干扰和异常, 本文对原始数据进行了系统的清洗和过滤。对于数据缺失区间少于 30 个采样点的情况, 采用前向填充方法进行补充; 对于缺失区间超过 30 个采样点的数据段, 则将其标记为无效数据并在后续分析中剔除。此外, 本文还对功率数据进行了有效范围限定: 设定了总功率和各电器功率的上限阈值, 同时将小于 5 W 的功率数据视为噪声并过滤。

3.3 模型评价指标

为全面评估所提模型的性能, 本文采用了 4 种评

价指标：平均绝对误差 (E_{MA})、平均相对误差 (E_{MR})、准确率 (A_c) 以及 F1 分数。各指标的计算公式如下。

(1) 平均绝对误差 E_{MA} 为

$$E_{MA} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (29)$$

(2) 平均相对误差 E_{MR} 为

$$E_{MR} = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{y_i} \quad (30)$$

式中： N 为样本数量； y_i 为电器 i 时刻的实际功率值； $\hat{y}$ 为模型预测的功率值。

(3) 准确率

$$A_c = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \quad (31)$$

式中： T_p 表示正确预测设备开启的样本数； T_n 表示正确预测设备关闭的样本数； F_p 表示错误预测设备开启的样本数； F_n 表示错误预测设备关闭的样本数。

(4) F1 分数

$$F1 = \frac{2PR}{P+R} \quad (32)$$

$$P = \frac{T_p}{T_p + F_p} \quad (33)$$

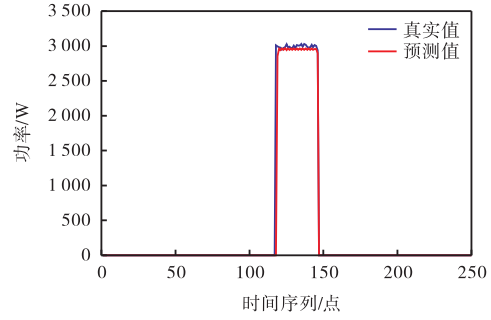
$$R = \frac{T_p}{T_p + F_n} \quad (34)$$

式中： P 为精确率，用于衡量模型预测为“开启”状态的样本中，实际确实是“开启”的比例； R 为召回率，用于衡量在所有实际为“开启”状态的样本中，被模型正确识别出来的比例；F1 分数是精确率和召回率的调和平均值，用于评估模型识别设备状态的综合性能。

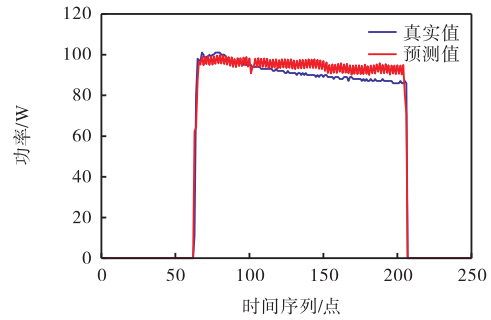
3.4 模型指标对比与数据可视化分析

图 4 呈现了模型在 UK-DALE 数据集上对水壶、冰箱、洗衣机、洗碗机 4 类电器的负荷分解效果。从图中可见，所提模型对于不同工作特性和功率水平的用电设备均表现出较好的识别效果，能够较为准确地刻画设备运行过程中的状态切换特征，并在正常工作区间维持稳定的预测水平，验证了所提模型在负荷分解中的可靠性与鲁棒性。

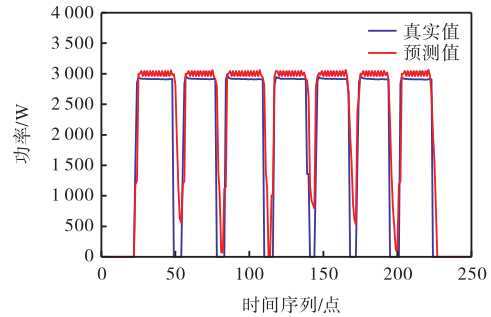
为验证所提出改进模型的有效性，本研究选取了正则化双向 GRU (R-BiGRU)、正则化双向 LSTM^[24]、(R-BiLSTM) 以及序列到序列模型 (Seq2seq)^[25] 作为对比对象。为确保实验的公平性，所有模型均采用相同的训练配置，并在未经训练的家用电器数据集上进行测试。表 3 和表 4 分别展示了各模型在 REDD 数据集和 UK-DALE 数据集上的性能指标。



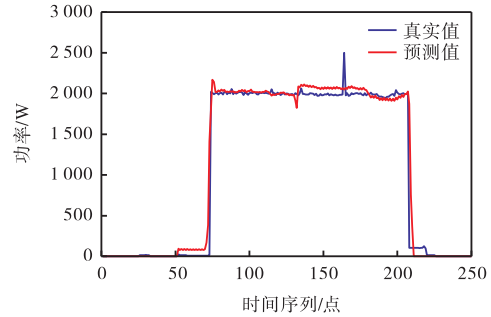
(a) 水壶负荷分解曲线



(b) 冰箱负荷分解曲线



(c) 洗衣机负荷分解曲线



(d) 洗碗机负荷分解曲线

图 4 改进模型负荷分解曲线

Fig.4 Load decomposition curves of improved model

从表 3 可以看出，所提模型在 REDD 数据集上的 3 类电器中取得了较好的表现。从整体来看，所提模型在 4 个评价指标上均优于其他对比模型。

从表 4 可以看出，对于 UK-DALE 数据集的 3 类电器设备，所提模型均取得了较好的结果。

表 3 REDD 性能指标对比

Tab.3 Comparison of REDD performance index

电器设备	模型	准确率	F1 分数	E_{MR}	E_{MA}
冰箱	R-BiGRU	0.794	0.705	0.829	44.28
	R-BiLSTM	0.798	0.709	0.841	44.82
	Seq2seq	0.796	0.689	0.822	35.69
	本文所提模型	0.888	0.813	0.820	32.55
洗衣机	R-BiGRU	0.922	0.216	0.090	27.63
	R-BiLSTM	0.989	0.125	0.020	35.73
	Seq2seq	0.970	0.274	0.042	36.12
	本文所提模型	0.993	0.723	0.014	18.38
洗碗机	R-BiGRU	0.955	0.034	0.042	25.29
	R-BiLSTM	0.956	0.421	0.056	25.25
	Seq2seq	0.953	0.298	0.053	25.29
	本文所提模型	0.944	0.480	0.035	25.41
平均值	R-BiGRU	0.890	0.318	0.320	32.40
	R-BiLSTM	0.914	0.418	0.306	35.27
	Seq2seq	0.906	0.418	0.306	32.37
	本文所提模型	0.942	0.672	0.290	25.45

表 4 UK-DALE 性能指标对比

Tab.4 Comparison of UK-DALE performance index

电器设备	模型	准确率	F1 分数	E_{MR}	E_{MA}
水壶	R-BiGRU	0.993	0.425	0.008	23.22
	R-BiLSTM	0.994	0.531	0.007	21.26
	Seq2seq	0.997	0.850	0.003	9.64
	本文所提模型	0.998	0.911	0.002	5.82
冰箱	R-BiGRU	0.636	0.401	0.901	39.54
	R-BiLSTM	0.573	0.174	0.956	43.74
	Seq2seq	0.772	0.718	0.758	29.20
	本文所提模型	0.849	0.820	0.683	20.95
洗衣机	R-BiGRU	0.342	0.018	0.062	68.65
	R-BiLSTM	0.938	0.150	0.067	15.66
	Seq2seq	0.913	0.173	0.094	11.90
	本文所提模型	0.945	0.305	0.060	10.51
洗碗机	R-BiGRU	0.977	0.639	0.035	38.42
	R-BiLSTM	0.976	0.605	0.033	36.36
	Seq2seq	0.947	0.560	0.069	25.43
	本文所提模型	0.981	0.723	0.034	20.03
平均值	R-BiGRU	0.737	0.371	0.252	42.46
	R-BiLSTM	0.870	0.365	0.266	29.25
	Seq2seq	0.907	0.575	0.231	19.04
	本文所提模型	0.943	0.690	0.195	14.33

从表 3 与表 4 的实验结果可见, 所提模型在不同数据集上对水壶、洗衣机及洗碗机 3 类电器设备均实现了较好的辨识准确率, 然而对于冰箱的辨识准确率在两个测试数据集中均略低于其他设备类型. 从图 4 呈现的 4 类电器负荷分解曲线分析可见, 水壶、洗衣机及洗碗机 3 类电器均具有较高的峰值运行功率, 其运行功率普遍超过 2 000 W, 在总负荷中占比显著. 因此, 在负荷分解过程中, 这些电器的运行状态变化会在总负荷曲线上产生明显的特征波动, 使模型能够有效提取并识别其特征信息.

相较而言, 冰箱在运行状态下的功率仅约 100 W, 仅为其他 3 类电器的约 5%. 由于其本身功率水平较低, 冰箱的开关状态变化对总负荷曲线的影响幅度有限, 这些微弱变化容易被其他电器的功率波动所掩盖或干扰, 导致特征信息被削弱. 因此, 模型对冰箱类负荷的识别准确率略低于其他电器.

3.5 联邦学习-知识蒸馏框架有效性验证

3.5.1 知识蒸馏有效性验证

为验证知识蒸馏技术在实现模型轻量化并提升轻量化模型性能方面的有效性, 本文基于 UK-DALE 数据集选取水壶和洗衣机两类电器进行实验. 实验中采用所提的 CNN-Transformer 作为教师模型, 轻量化 LSTM 作为学生模型, 并额外加入 R-BiLSTM 作为对比模型. 知识蒸馏过程中的温度参数 T 和超参数 α 分别设置为 3.0 和 0.5.

本文从两个维度验证知识蒸馏的有效性. 第 1 个维度基于所提联邦学习-知识蒸馏框架, 云端仅存储 3、4 号房屋的公共数据集. 基于框架运行状态, 首先利用该数据对学生模型进行无知识蒸馏训练获得基准模型; 随后在相同数据集上, 通过预训练教师模型进行知识蒸馏训练, 得到知识蒸馏学生模型.

第 2 个维度验证知识蒸馏在少样本场景下的特性. 设计多样本量和少样本量两组对照实验: 多样本量组采用 1、3、4、5 号房间的完整数据集; 少样本量组仅使用 3、4 号房屋数据.

本文框架中教师模型和 R-BiLSTM 对照模型均采用 1、3、4、5 号房屋的完整数据集训练. 在知识蒸馏阶段, 教师模型仅使用云端存储的 3、4 号房屋数据指导学生模型, 以符合框架实际部署场景.

从表 5 与表 6 模型性能对比结果可见, 针对水壶负载, 在少样本训练场景下, 采用知识蒸馏的学生模型在准确率、F1 分数、 E_{MR} 和 E_{MA} 4 个评估指标上均优于无知识蒸馏的学生模型. 同时, 经过知识蒸馏的学生模型性能接近于使用多样本训练的对比模型 R-BiLSTM.

对于洗衣机负载, 知识蒸馏同样展现出显著的性能提升效果. 在少样本训练场景下, 知识蒸馏不仅提升了学生模型在各项评估指标上的表现, 在误差指标

表 5 基于水壶的性能对比

Tab.5 Comparison of kettle-based performance

模型	知识蒸馏	样本量	准确率	F1 分数	E_{MR}	E_{MA}
教师模型	无	多	0.998	0.911	0.002	5.82
学生模型	无	多	0.996	0.736	0.006	16.74
	无	少	0.993	0.357	0.008	24.13
	有	少	0.996	0.703	0.006	16.53
R-BiLSTM	无	多	0.994	0.531	0.007	21.26

表 6 基于洗衣机的性能对比

Tab.6 Comparison of washing machine-based performance

模型	知识蒸馏	样本量	准确率	F1 分数	E_{MR}	E_{MA}
教师模型	无	多	0.945	0.305	0.060	10.51
学生模型	无	多	0.805	0.115	0.205	22.40
	无	少	0.775	0.104	0.234	24.06
	有	少	0.950	0.162	0.055	16.10
R-BiLSTM	无	多	0.938	0.150	0.067	15.66

方面优于采用多样本训练的 R-BiLSTM 对比模型。

除了模型性能精度之外,为全面评估所提框架的实用性,本文进一步对比了教师模型、学生模型和 R-BiLSTM 对比模型在模型参数量、模型文件大小、训练时间和推理时间 4 个方面的性能指标。其中,训练时间基于 UK-DALE 数据集中 1、3、4、5 号房屋的水壶负载数据测得;推理时间则基于 2 号房屋水壶负载数据测得。

从表 7 可见,学生模型展现显著的轻量化效果。模型参数量为 20 万,相较于教师模型和 R-BiLSTM 分别减少了 90%和 84.6%;存储空间仅需 1.64 MB,较两者分别减少 89%和 83%。训练时间仅需 2 min 56 s,较两者分别缩短 70%和 74%;推理用时 8 s,分别减少 43%和 47%。

表 7 模型参数量与时间对比

Tab.7 Comparison of model parameters and time

模型	参数量/万	文件大小/MB	训练时间	推理时间/s
教师模型	200	14.80	9 min 51 s	14
学生模型	20	1.64	2 min 56 s	8
R-BiLSTM	130	9.67	11 min 10 s	15

综上所述,从表 5~7 的对比分析可见,知识蒸馏在轻量化与性能提升两方面均显著。对于水壶和洗衣机负载,学生模型在各评估指标上均有明显提升,且在少样本训练场景下可达到甚至超过多样本训练的 R-BiLSTM 水平。在轻量化方面,学生模型仅需约 10%的参数量和存储空间,训练和推理时间显著降低。实验表明,知识蒸馏通过云端知识迁移,在保持性能的同时降低了模型复杂度,为边缘侧轻量化部署和联邦学习分布式训练提供了有效方案。

3.5.2 联邦学习有效性验证

为验证所提联邦学习框架的有效性,本文构建了基于云边协同的分布式训练验证实验。实验包含一个云端节点和两个边缘节点,云端配置 3、4 号房屋数据作为公共数据集,边缘节点 1 和 2 分别配置 1 号和 5 号房屋数据作为私有数据集。两个边缘节点采用相同的训练配置,所有模型均采用 2 号房屋数据作为统一测试集。

本文采用基于知识蒸馏的轻量化学生模型作为初始全局模型,设计一轮联邦学习迭代验证。训练时,首先将初始全局模型(全局模型 1)参数分发至边缘节点,各节点基于全局模型 1 和本地私有数据进行训练与测试,随后上传模型参数至云端。云端通过联邦聚合算法融合参数生成新的全局模型(全局模型 2)并验证性能。在整个过程中,边缘节点的私有数据始终保持在本地图,云边节点间仅传输模型参数,实现了隐私保护下的分布式训练与优化。

从表 8 可见,对于水壶设备,边缘节点 1、2 的模型性能相较于全局模型 1 均实现显著提升。经联邦聚合后,全局模型 2 继承了边缘训练的性能优势,其 4 个评估指标均明显改善。

表 8 联邦学习模型指标

Tab.8 Index of federated learning model

电器设备	模型	准确率	F1 分数	E_{MR}	E_{MA}
水壶	全局模型 1	0.996	0.703	0.006	16.533
	全局模型 2	0.998	0.863	0.004	10.366
	教师模型	0.998	0.911	0.002	5.820
	边缘节点 1	0.999	0.929	0.003	9.120
	边缘节点 2	0.997	0.848	0.003	9.553
洗衣机	全局模型 1	0.949	0.162	0.055	16.095
	全局模型 2	0.939	0.218	0.067	27.580
	教师模型	0.945	0.305	0.063	19.510
	边缘节点 1	0.955	0.198	0.05	22.228
	边缘节点 2	0.936	0.220	0.072	16.459

对于洗衣机设备,全局模型 1 经边缘节点 1 训练后性能显著提升;但在边缘节点 2,仅 F1 分数获得提升,其余指标均有所下降。分析表明,这种性能衰减主要源于 5 号房间洗衣机数据质量较低。经联邦聚合后,全局模型 2 的 F1 分数显著提升,但准确率和 E_{MR} 小幅下降, E_{MA} 降低较大,说明低质量边缘数据会影响全局模型的聚合效果。

综上所述,所提框架实现了隐私保护下的分布式训练。通过云边协同机制,边缘节点能利用私有数据优化模型,且优化效果可通过联邦聚合迁移至全局模型。对水壶负载,框架展现优异的性能提升;对洗衣机负载,虽受数据质量影响,但仍体现出优化效果。

3.5.3 不同框架的通信数据量对比

为定量评估框架的通信优势,本文对 3 种框架进行对比:传统分布式机器学习框架、常规联邦学习框架及所提的联邦学习-知识蒸馏框架。

传统分布式框架需将边缘节点的原始数据传输至云端,导致显著通信开销;常规联邦学习框架通过只传输模型参数降低通信开销,但当模型参数量大时仍产生较大负载。本文所提框架采用轻量级学生模型作为全局模型,由于其参数规模显著减小,能大幅

降低云边协同的通信开销。

为了量化对比 3 种框架的通信流量,本实验采用与前述联邦学习验证实验相同的云边部署配置,在单轮迭代场景下进行测试。

表 9 展示了两个边缘节点的数据集详情。选取水壶和洗衣机作为测试电器,由于非侵入式负荷分解需同时获取单个电器的分类负荷和家庭总用电量数据,表中列出了两个节点的总表数据、水壶用电数据和洗衣机用电数据的容量大小。

表 10 对比了不同框架的通信消耗。传统分布式学习框架的上行通信量为两节点数据集大小之和,下行通信量为模型文件大小乘以节点数,总通信量为上、下行通信量之和。联邦学习框架因无需传输原始

数据,通信量主要来自模型参数传输。每轮迭代中,节点需上传本地模型参数并接收更新后的全局模型,因此上、下行通信量相等,均为模型文件大小乘以节点数,总通信量为上、下行通信量之和。

表 9 边缘节点数据量

Tab.9 Data volume at edge nodes

节点	数据集类型	数据集大小/MB
边缘节点 1	总表数据	313.0
	水壶数据	234.0
	洗衣机数据	244.0
边缘节点 2	总表数据	25.3
	水壶数据	12.2
	洗衣机数据	24.6

表 10 不同框架的通信量对比

Tab.10 Communication traffic comparison among different frameworks

所用框架	所用模型	模型大小/MB	上行通信量/MB	下行通信量/MB	总通信量/MB
传统分布式学习	R-BiLSTM	9.67	853.10	9.67	862.77
联邦学习	教师模型	14.80	29.60	29.60	59.20
联邦学习	R-BiLSTM	9.67	19.34	19.34	38.68
本文框架	学生模型	1.64	3.28	3.28	6.56

从表 9 数据可见,3 种框架通信开销差异显著。传统分布式学习框架需传输大量原始数据,总通信量达 862.77 MB。常规联邦学习框架仅传输模型参数,采用教师模型和 R-BiLSTM 时,总通信量分别为 59.20 MB 和 38.68 MB。本文框架通过引入轻量级学生模型,将通信量降至 6.56 MB。与传统框架相比降低 99.24%;较采用教师模型和 R-BiLSTM 的联邦框架分别减少 88.92% 和 83.04%。

分布式框架的通信开销与系统规模和训练过程强相关。随节点数量增加和训练轮次延长,通信量呈线性增长。传统框架中,新增节点导致更多原始数据上传;联邦框架中,每轮迭代需与所有节点交换参数。在数据规模不变时,理论分析表明本文框架需约 130 轮迭代才能达到传统框架的通信开销。此外,本文框架与传统联邦框架采用相同迭代范式,但通过轻量级模型显著降低了各轮次通信开销,验证了其在降低分布式训练通信开销方面的有效性。

4 结 论

针对深度学习模型在非侵入式负荷分解中面临的数据隐私保护和边缘部署效率问题,本文提出了一种基于联邦学习与知识蒸馏的轻量化负荷分解框架,通过实验验证得出以下结论。

(1) 采用 CNN-Transformer 混合架构构建了一种改进的负荷分解模型,利用 CNN 提取负荷序列的局部特征,通过 Transformer 增强长期时序依赖建模能力。所提模型在 REDD 和 UK-DALE 两个公开数据集的负荷分解任务中展现出更优的辨识性能。

(2) 提出联邦学习-知识蒸馏框架,实现了模型的高效轻量化与隐私保护。通过知识蒸馏技术,轻量化学生模型在保持了较好的辨识性能的情况下,模型的参数规模降低了 90%,训练时间减少了 70%,推理速度提升了 43%。同时引入联邦学习,使得边缘节点无需共享原始数据即可参与模型训练,在保护用户隐私的同时有效提升了模型性能。

(3) 相比传统分布式学习和常规联邦学习方法,所提框架引入轻量化模型大幅降低了云边协同过程中的通信开销,相比传统框架通信量降低了 99.24%,较采用教师模型的传统联邦学习框架通信量降低了 88.92%,轻量化模型在训练和推理阶段表现出更高的计算效率,探索了模型在边缘计算等资源受限场景下的部署方案。

参考文献:

- [1] 于 浩, 杨家辉, 习 伟, 等. 智能配用电数据压缩方法研究进展及展望[J]. 供用电, 2024, 41(8): 2-14.

- Yu Hao, Yang Jiahui, Xi Wei, et al. Research progress and prospects on compression methods of smart power distribution and consumption data[J]. Power Supply and Consumption, 2024, 41(8): 2-14(in Chinese).
- [2] Tanoni G, Stankovic L, Stankovic V, et al. Knowledge distillation for scalable nonintrusive load monitoring[J]. IEEE Transactions on Industrial Informatics, 2024, 20(3): 4710-4721.
- [3] 郭陆阳, 王守相, 陈海文, 等. 基于双向 GRNN 与时间序列翻译模型的非侵入式负荷分解算法[J]. 供用电, 2019, 36(10): 9-15, 86.
- Guo Luyang, Wang Shouxian, Chen Haiwen, et al. Non-invasive load decomposition algorithm based on bidirectional GRNN and time series translation model[J]. Power Supply and Consumption, 2019, 36(10): 9-15, 86(in Chinese).
- [4] 冷喜武, 陈国平, 蒋 宇, 等. 智能电网监控运行大数据分析系统的数据规范和数据处理[J]. 电力系统自动化, 2018, 42(19): 169-176.
- Leng Xiwu, Chen Guoping, Jiang Yu, et al. Data specification and processing in big-data analysis system for monitoring and operation of smart grid[J]. Automation of Electric Power Systems, 2018, 42(19): 169-176(in Chinese).
- [5] Hart G W. Nonintrusive appliance load monitoring[J]. Proceedings of the IEEE, 1992, 80(12): 1870-1891.
- [6] 全 杰, 齐子豪, 蒲天骄, 等. 电力物联网边缘智能: 概念、架构、技术及应用[J]. 中国电机工程学报, 2024, 44(14): 5473-5496.
- Tong Jie, Qi Zihao, Pu Tianjiao, et al. Edge intelligence to power internet of things: Concept, architecture, technology and application[J]. Proceedings of the CSEE, 2024, 44(14): 5473-5496(in Chinese).
- [7] Schirmer P A, Mporas I. Non-intrusive load monitoring: A review[J]. IEEE Transactions on Smart Grid, 2022, 14(1): 769-784.
- [8] Schirmer P A, Mporas I, Paraskevas M. Energy disaggregation using elastic matching algorithms[J]. Entropy, 2020, 22(1): 71.
- [9] Zhong M, Goddard N, Sutton C. Interleaved factorial non-homogeneous hidden Markov models for energy disaggregation[EB/OL]. <https://arxiv.org/abs/1406.7665>, 2014-06-30.
- [10] Kelly J, Knottenbelt W. Neural NILM: Deep neural networks applied to energy disaggregation[C]//Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments. Seoul, Republic of Korea, 2015: 55-64.
- [11] Yue Z, Witzig C R, Jorde D, et al. BERT4NILM: A bidirectional transformer model for non-intrusive load monitoring[C]//Proceedings of the 5th International Workshop on Non-Intrusive Load Monitoring. Japan, 2020: 89-93.
- [12] Bao K, Ibrahimov K, Wagner M, et al. Enhancing neural non-intrusive load monitoring with generative adversarial networks[J]. Energy Informatics, 2018, 1: 295-302.
- [13] 鲍光海, 黄逸欣. 基于 ResNeXt 网络和迁移学习的非侵入式负荷监测[J]. 电力系统自动化, 2023, 47(13): 110-120.
- Bao Guanghai, Huang Yixin. Non-intrusive load monitoring based on ResNeXt network and transfer learning[J]. Automation of Electric Power Systems, 2023, 47(13): 110-120(in Chinese).
- [14] Pan Y, Liu K, Shen Z, et al. Sequence-to-subsequence learning with conditional gan for power disaggregation[C]//2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Barcelona, Spain, 2020: 3202-3206.
- [15] 何耿生, 黄 宇, 曾金灿, 等. 基于多通道融合机制的非侵入式负荷监测方法[J]. 计算机应用, 2025, 45(增 1): 317-322.
- He Gengsheng, Huang Yu, Zeng Jincan, et al. Non-intrusive load monitoring method based on multi-channel fusion mechanism[J]. Computer Applications, 2025, 45(Suppl1): 317-322(in Chinese).
- [16] 王守相, 郭陆阳, 陈海文, 等. 基于特征融合与深度学习的非侵入式负荷辨识算法[J]. 电力系统自动化, 2020, 44(9): 103-110.
- Wang Shouxian, Guo Luyang, Chen Haiwen, et al. Non-intrusive load identification algorithm based on feature fusion and deep learning[J]. Automation of Electric Power Systems, 2020, 44(9): 103-110(in Chinese).
- [17] 凡 航, 徐 葳, 范晓昱, 等. 隐私计算在新型电力系统中的应用分析与展望[J]. 电力系统自动化, 2023, 47(19): 187-199.
- Fan Hang, Xu Wei, Fan Xiaoyu, et al. Application analysis and prospect of privacy-preserving computation in new power system[J]. Automation of Electric Power Systems, 2023, 47(19): 187-199(in Chinese).
- [18] 宋祺鹏, 王继东, 张丽伟, 等. 本地化差分隐私下的

- 电力物联网终端数据隐私保护方法[J]. 重庆邮电大学学报(自然科学版), 2023, 35(6): 1001-1010.
- Song Qipeng, Wang Jidong, Zhang Liwei, et al. Terminal data privacy protection method for power Internet of Things based on localized differential privacy[J]. Journal of Chongqing University of Posts and Telecommunications(Natural Science Edition), 2023, 35(6): 1001-1010(in Chinese).
- [19] Kukunuri R, Aglawe A, Chauhan J, et al. EdgeNILM: Towards NILM on edge devices[C]//Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation. Japan, 2020: 90-99.
- [20] McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. Florida, USA, 2017: 1273-1282.
- [21] Dai S, Meng F, Wang Q, et al. DP2-NILM: A distributed and privacy-preserving framework for non-intrusive load monitoring[J]. Renewable and Sustainable Energy Reviews, 2024, 191: 114091.
- [22] Lin Y H, Ciou J C. A privacy-preserving distributed energy management framework based on vertical federated learning-based smart data cleaning for smart home electricity data[J]. Internet of Things, 2024, 26: 101222.
- [23] Hinton G. Distilling the knowledge in a neural network[EB/OL]. <https://arxiv.org/pdf/1503.02531>, 2015-03-09.
- [24] Rafiq H, Zhang H X, Li H M, et al. Regularized LSTM based deep learning model: First step towards real-time non-intrusive load monitoring[C]//2018 IEEE International Conference on Smart Energy Grid Engineering(SEGE). Oshawa, Canada, 2018: 234-239.
- [25] Zhang C Y, Zhong M J, Wang Z Z, et al. Sequence-to-point learning with neural networks for non-intrusive load monitoring[C]//Proceedings of the AAAI Conference on Artificial Intelligence. New Orleans, USA, 2018: 2604-2611.

(责任编辑: 孙立华)