

文章编号:1672-3961(2024)03-0055-09 DOI:10.6040/j.issn.1672-3961.0.2023.217

基于联邦学习的时间序列预测算法

刘新^{1,2}, 刘冬兰^{1,2,*}, 付婷³, 王勇⁴, 常英贤⁴, 姚洪磊^{1,2}, 罗昕³, 王睿^{1,2}, 张昊^{1,2}

(1. 国网山东省电力公司电力科学研究院, 山东 济南 250003; 2. 山东省智能电网技术创新中心, 山东 济南 250003; 3. 山东大学软件学院, 山东 济南 250101; 4. 国网山东省电力公司, 山东 济南 250001)

摘要: 为应对不断升级的数据隐私保护需求, 提出一种基于分布式场景下的时间序列预测算法。该算法主要改进体现在以下两个方面: 在客户端模型本地训练阶段, 通过正则化项约束本地模型训练方向, 解决本地模型漂移问题; 在全局模型聚合阶段, 提出客户端贡献估计策略, 根据客户端贡献程度分配权重, 保护客户端协作公平性, 提升全局模型泛化能力。为验证改进后算法有效性, 在 ETTh1 数据集、ETTm1 数据集和 Weather 数据集上将其与基线联邦学习算法 FedAvg 对比。试验结果表明, 改进后算法在 ETTh1 数据集上均方误差 E_{MS} 平均降低 2.99%, 在 ETTm1 数据集上 E_{MS} 平均降低 3.57%。在算法中加入正则化项和客户端贡献估计策略, E_{MS} 分别下降 0.84% 和 2.78%, 同时加入这两个模块, E_{MS} 降低 3.03%, 验证提出的算法在预测性能方面表现出更高预测准确性。

关键词: 联邦学习; 机器学习; 时间序列预测; 分布式系统; 深度学习**中图分类号:** TP181 **文献标志码:** A**引用格式:** 刘新, 刘冬兰, 付婷, 等. 基于联邦学习的时间序列预测算法[J]. 山东大学学报(工学版), 2024, 54(3): 55-63.

LIU Xin, LIU Donglan, FU Ting, et al. Time series forecasting algorithm based on federated learning[J]. Journal of Shandong University (Engineering Science), 2024, 54(3): 55-63.

Time series forecasting algorithm based on federated learning

LIU Xin^{1,2}, LIU Donglan^{1,2,*}, FU Ting³, WANG Yong⁴, CHANG Yingxian⁴, YAO Honglei^{1,2}, LUO Xin³, WANG Rui^{1,2}, ZHANG Hao^{1,2}

(1. State Grid Shandong Electric Power Research Institute, Jinan 250003, Shandong, China; 2. Shandong Smart Grid Technology Innovation Center, Jinan 250003, Shandong, China; 3. School of Software, Shandong University, Jinan 250101, Shandong, China; 4. State Grid Shandong Electric Power Company, Jinan 250001, Shandong, China)

Abstract: To meet the constantly evolving demands for data privacy protection, a time series forecasting algorithm designed for distributed scenarios was proposed. The main improvements of the algorithm were as follows: during the local model training phase at client nodes, a regularization term was employed to constrain the training direction of local model, resolving the issue of local model drift; in the global model aggregation phase, a client contribution estimation strategy was proposed, which allocated weights based on the extent of client contributions to ensure fairness in client collaboration and enhance the generalizability of global model. To validate the effectiveness of improved algorithm, it was compared with the baseline federated learning algorithm FedAvg on ETTh1, ETTm1, and Weather datasets. Experimental results showed that the improved algorithm reduced the mean square error E_{MS} by 2.99% on ETTh1 dataset and 3.57% on ETTm1 dataset. Incorporating the regularization term and client contribution estimation strategy into the algorithm led to respective decreases of 0.84% and 2.78% in E_{MS} , and a combined decrease of 3.03% in E_{MS} , confirming that the proposed algorithm showed higher predictive accuracy.

Keywords: federated learning; machine learning; time series forecasting; distributed systems; deep learning

收稿日期: 2023-09-05

基金项目: 国网山东省电力公司科技资助项目(520626220018)

第一作者简介: 刘新(1981—), 女, 山东枣庄人, 高级工程师, 硕士, 主要研究方向为网络安全、数据安全、隐私计算、区块链等。

E-mail: 49557599@qq.com

* 通信作者简介: 刘冬兰(1987—), 女, 云南宣威人, 高级工程师, 硕士, 主要研究方向为网络安全、数据安全、隐私计算、区块链等。

E-mail: liudonglan2006@126.com

0 引言

时间序列预测作为一种数据分析技术,通过对过去时间序列数据进行分析,完成对未来时间点数据的预测^[1-2]。在实时规划和资源分配方面,时间序列预测扮演重要角色,广泛应用于能源需求、交通流量和天气预测等领域^[3-5]。日益加强的法律限制和隐私安全保护要求,使传统中心化训练方法变得不切实际^[6]。为解决这个问题,联邦学习应运而生。作为一种新兴的分布式机器学习范式,联邦学习能保护数据隐私,避免大规模数据传输,降低网络带宽要求,增强模型训练的灵活度并提升效率^[7-8]。使用基于联邦学习的方式训练时间序列预测模型主要面临以下两方面挑战^[9]:(1)在客户端数据异质性情况下,不同客户端间数据分布存在差异,影响全局模型收敛性^[10];(2)现有根据客户端数据量来衡量各本地模型重要性的方法缺乏公平性,导致全局模型在数据量较少,具有独特数据的客户端上表现较差^[11]。

为解决上述问题,本研究提出一种基于联邦学习的时间序列预测算法。引入一个正则化项缓解真实应用场景下各客户端间数据异质性挑战,提高本地模型通用性。利用客户端贡献估计策略,动态调整不同客户端的贡献程度,解决联邦学习中不公平性带来的挑战,提高全局模型性能和鲁棒性,应对广泛存在的隐私保护问题。

1 算法可行性研究

时间序列预测算法以其卓越预测能力和实时应用优势,有效实现资源合理调配和规划,提高效率和资源利用率。在当前技术环境下,大部分时间序列预测算法依赖大型数据集。联邦学习使时间序列预测算法基于不同设备或机构数据训练,促进算法从多样化数据源中获取知识,提高模型泛化能力。将联邦学习与时间序列预测算法结合,在保护数据隐私同时显著提高时间序列预测算法效果,促进双方合作共赢,展现广阔研究前景。

1.1 时间序列预测

在时间序列预测领域,已经涌现出众多相关研究。传统统计学方法包括移动平均法、指数平滑法和 ARIMA 模型^[12-13]。这类经典方法往往仅适用于平稳和简单的时间序列数据,面对复杂序列表现不佳。随着深度学习的迅速发展,神经网络在时间序

列预测领域展现出显著成果,深度学习方法能突破传统单序列建模的限制,捕捉复杂时间序列中的内在规律^[14]。在时间序列预测领域,深度学习模型主要包括循环神经网络^[15](recurrent neural network, RNN)、卷积神经网络^[16](convolutional neural network, CNN)和 Transformers^[17]等。这些网络结构能有效捕捉时间序列数据中的长期依赖关系,在处理复杂非线性时间序列数据时表现出色。DeepAR 采用经典的长短期记忆网络(long-short term memory, LSTM)结构进行未来数据预测^[18]。基于 CNN 的时间序列预测模型则采用空洞卷积与因果卷积的结合网络,保持时间复杂度的同时引入更长的历史序列信息,提升预测效果。Transformers 引入自注意力机制,取代传统 RNN 和 CNN 等序列建模方法,更好捕捉时间序列中的长期依赖关系^[19-20]。Informer 引入 ProbSparse 自注意力机制,提升 Transformer 在长周期预测中的运行效率^[21]。这些深度学习模型为时间序列预测领域带来新的可能性,提高预测的准确性和效率,使其能更好应对复杂的实际问题。

1.2 联邦学习

作为联邦学习的经典算法之一,FedAvg 提供一种简单而实用的方法来解决模型聚合问题^[22]。但 FedAvg 的性能难以得到充分保障。为解决上述问题,一些算法对 FedAvg 进行改进。FedProx 在本地目标中引入近端项来约束本地更新方向,以实现本地和全局优化目标的协调^[23]。MOON 则利用对比学习来对齐本地模型更新和全局模型,以修正本地模型漂移^[24]。FedAlign 尝试利用正则化方法提升本地学习的通用性来解决数据异质性问题^[25]。但 these 方法仅通过简单的模型聚合来获得全局模型,忽略本地知识不兼容性,导致全局模型的泛化能力和收敛性难以得到保证^[26-27]。

2 问题研究

在联邦学习范式中,拥有数据的设备或机构通常被称为客户端,进行全局处理的设备或机构称为中央服务器。联邦学习能够保证各个客户端不共享私有数据的前提下,通过中央服务器协调各客户端在其本地使用私有数据进行本地模型训练,将本地模型参数发送到中央服务器进行模型聚合,获得一个可共享的全局模型^[28-29]。

2.1 问题定义

本研究提出一种全新的基于联邦学习的时间

序列预测算法。在联邦学习场景下,假设存在 K 个客户端,第 k 个客户端的本地数据表示为 D_k ,算法目标是通过中央服务器的协调,获得一个全局模型,同时确保各客户端之间不共享其私有数据。传统联邦学习的总体目标可以表述为:

$$\min_w f(w) = \sum_{k=1}^K \frac{|D_k|}{|D|} f_k(w), \quad (1)$$

式中, w 是全局模型参数, $f(w)$ 是全局目标, $|D_k|$ 是第 k 个客户端中本地样本数量, $|D| = \sum_{k=1}^K |D_k|$ 代表所有客户端的样本总数, $f_k(w)$ 是第 k 个客户端的损失函数。通过在分布式客户端上进行协作训练,不仅能保证客户端数据安全,还能获得一个能最小化所有客户端训练损失的全局模型,以保证全局模型的性能和泛化能力。

2.2 本地漂移约束

尽管联邦学习可以取代传统集中式机器学习范式,降低隐私泄露风险并保护客户端数据隐私,它也面临一些挑战。挑战之一是数据异构性,即客户端的数据可能是非独立同分布的。客户端间数据异构性可能会导致各客户端在本地数据集上训练的本地模型与在全局数据集上训练的全局模型间存在差异,使客户端本地模型朝着与全局模型不一致的方向进行优化,引发优化方向偏差,甚至灾难性遗忘全局知识。尽管在每个客户端本地训练过程中,本地模型能收敛到局部最优解,由于各客户端优化方向不一致,可能会使得全局模型的收敛性存在不确定性。

为缓解本地模型漂移,本研究引入一个动态正则化项,通过引入漂移变量减少本地漂移对全局模型优化的影响。

在本地训练阶段,本研究提出的算法为每个客户端定义一个本地漂移变量 v_k ,用以衡量本地模型与全局模型间差异,表示为 $v_k = w - w_k$,其中 w_k 为第 k 个客户端中本地模型参数, w 为全局模型参数。为在本地训练中减少本地模型参数和全局模型参数间差距,设计如下形式的正则化项:

$$R_k(w_k, v_k, w) = \|v_k + w_k - w\|^2, \quad (2)$$

式中, R_k 表示第 k 个客户端的正则化项, v_k 为第 k 个客户端的漂移变量。在本地训练过程中,正则化项 R_k 越大,代表客户端 k 的本地模型漂移程度越高。客户端根据漂移变量对本地模型训练进行约束,不仅能减少本地训练过程中模型漂移的负面影响,还能约束本地模型优化方向,提高全局模型的鲁棒性。客户端的总体目标函数表示为如下形式:

$f_k(w_k) = \mathcal{L}_k((x_k, y_k); w_k) + \eta R_k(w_k, v_k, w)$, (3) 式中: \mathcal{L}_k 为本地经验损失,可以采用时间序列预测模型中常见的损失函数; $(x_k, y_k) \in D_k$, 代表客户端 k 的本地数据; η 为超参数,用来控制正则化项的权重。

2.3 客户端公平性保证

在现有联邦学习算法中,中央服务器通常通过加权平均各客户端的模型更新来更新全局模型,协同寻找最优的全局模型参数。某个客户端的模型权重通常与其样本数量成正比,这种权重分配方式的直观解释是,本地数据量较大的客户端比数据量较小的客户端在优化过程中表现更好,因此赋予它们更大权重。在客户端数据异质性情况下,数据量较大的客户端并不一定对全局模型性能产生重大贡献。当某些客户端拥有与其他客户端不同的样本时,来自这些客户端的信息可能更有助于提升全局模型的性能。

本研究致力于优化中央服务器上的模型聚合过程,提升全局模型的鲁棒性和公平性。

在第 t 轮通信过程中,当参与训练的所有客户端完成本地训练后,计算全局模型参数和客户端 k 本地模型参数间的梯度,表示为 $\nabla G_k^t(w_k^t)$, $\nabla G^t(w^t) = \sum_{k=1}^K \alpha_k^{t-1} \nabla G_k^t(w_k^t)$ 表示全局梯度,其中 α_k^{t-1} 为第 $(t-1)$ 轮客户端 k 的本地模型权重, K 为客户端总数。除客户端 k 外,聚合的模型梯度可表示为如下形式:

$$\nabla G^t(w_{-k}^t) = \nabla G^t(w^t) - \alpha_k^{t-1} \nabla G_k^t(w_k^t). \quad (4)$$

客户端 k 与其他客户端的梯度间相似性越低,表示客户端 k 与其他客户端的优化方向越不同。通过捕捉这种差异性,可以帮助训练一个更具有泛化能力的全局模型。采用余弦相似性来度量两者间的相似性,具体公式为:

$$\alpha_k^t = \frac{1 - \cos(\nabla G_k(w_k), \nabla G(w^{-k}))}{\sum_{k=1}^K (1 - \cos(\nabla G_k(w_k), \nabla G(w^{-k})))}, \quad (5)$$

式中: α_k^t 代表客户端 k 在第 t 轮更新时的模型权重; $\cos(\cdot)$ 代表余弦相似度,当客户端 k 与其余客户端间余弦相似度越低时,表示该客户端代表新的优化方向信息。本研究为该客户端分配更大的权重。通过在梯度空间度量各客户端的重要程度,能保证联邦学习的公平性,最大限度利用来自客户端本地模型的知识,提高全局模型的鲁棒性和收敛速度。

利用新的权重来聚合本地模型参数,以得到全局模型:

$$w^{t+1} = \sum_{k=1}^K \alpha_k^t w_k^t, \quad (6)$$

式中, w^{t+1} 表示第 $(t+1)$ 轮的全局模型参数。这种简单而高效的重新加权聚合机制能够获得一个更具鲁棒性的全局模型, 确保全局模型充分捕捉(融合)各客户端的特征和知识。

2.4 算法概述

本研究提出一种基于联邦学习的时间序列预测算法, 旨在解决分布式场景下的协同优化问题。该问题的解决方案可以通过交替完成以下两个步骤来实现: (1) 在客户端本地训练阶段, 引入正则化项来限制本地模型漂移; (2) 在服务器端全局模型聚合阶段, 通过在梯度空间中度量客户端的重要程度, 确保联邦学习的公平性, 增强全局模型泛化性能。有关这两个步骤的详细信息, 请参阅算法 1。

算法 1 基于联邦学习的时间序列预测算法

输入 通信轮数 T ; 客户端数量 K ; 客户端本地数据集 $\{D_k\}_{k \in \{1, \dots, K\}}$; 本地训练轮数 E ; 初始模型参数 w^0 ,

输出 第 $(t+1)$ 轮的全局模型参数 w^{t+1} 。

服务器执行:

- (1) 在第 t 轮通信中, 将全局模型参数 w^t 分发给 K 个客户端;
- (2) 各客户端执行本地更新过程, 得到客户端更新的模型参数 w_k^t ;
- (3) 计算客户端 k 的模型梯度以及全局梯度;
- (4) 使用式(5)计算客户端 k 在第 t 轮的模型权重 α_k^t ;
- (5) 使用式(6)更新第 $(t+1)$ 轮的全局模型参数 w^{t+1} 。

客户端本地更新:

- (1) 从服务器端下载 w^t ;
- (2) 计算客户端 k 的本地模型参数 w_k^{t-1} 和全局模型参数 w^t 间的本地漂移变量 v_k ;
- (3) 对于设定的本地训练轮数 E , 重复执行以下步骤(4)–(6);
- (4) 通过式(2)使用本地漂移变量计算正则化损失;
- (5) 通过式(3)计算客户端总体损失;
- (6) 根据损失函数的值更新本地模型;
- (7) 返回客户端模型参数 w_k^t 。

3 试验

3.1 数据集

本研究在两个广泛用于时间序列预测任务的

基准数据集上进行试验。试验结果清楚表明, 该算法在全局模型性能上优于现有的基线方法。

ETT^[30]: 此数据集包含从电力变压器中收集的数据, 它代表电力系统长期部署的关键性能指标。其中 ETTh1 数据集包含从 2016—2018 年每小时记录的数据, 每个数据点包含 6 个负载指标以及与之相关的目标值, 即变压器油温。ETTm1 数据集涵盖 2016—2018 年, 以每 15 min 为间隔记录的变压器性能指标。为评估算法性能, 本研究将这两个数据集都按照 8:2 的比例划分为训练集和测试集。

Weather: 该数据集涵盖 2020 年全年的天气数据, 每隔 10 min 记录一次。每个数据点包括 21 项气象指标, 如空气湿度、温度和风速等, 提供全面的气象信息。在研究中按照标准协议, 将整个数据集按照 8:2 的比例划分为训练集和测试集, 以支持算法性能的评估。

3.2 试验细节

对于每个数据集, 均选用 Informer 作为时间序列预测模型的基础网络架构, 用于进行性能评估。为验证提出算法是否有效, 本研究选取 3 种最先进的联邦学习算法进行对比分析, 分别是 FedAvg、FedProx 和 MOON。FedAvg 是传统的联邦学习算法; FedProx 在客户端本地学习中引入一个近端项, 以约束本地模型更新更接近于全局模型; MOON 则通过模型表示间相似性来纠正本地模型漂移。为获得联邦学习领域中算法性能的上界, 采用集中式 (Centralized) 方式将数据集中起来训练模型作为对比。所有方法的主干网络都相同, 且从头开始训练, 不加载任何预训练模型权重。为确保公平评估, 在测试集上使用全局模型进行测试, 对所有试验进行 3 次重复以获得平均结果。

3.3 实现和超参数

在每个客户端上, 使用 Adam 优化器进行本地训练。本地训练的轮数设置为 1, batch size 设置为 32, 总通信轮次设置为 20, 客户端数量设置为 5, 学习率初始值为 0.000 1。在前述先进联邦学习算法 FedProx 和 MOON 中, 都存在一个超参数用于平衡额外的损失项。对于所有数据集, FedProx 和 MOON 的最优超参数 μ 分别为 1.0 和 0.1。对于本研究提出的算法, 在 ETTh1 数据集、ETTm1 数据集和 Weather 数据集上, 最优的超参数 η 分别为 0.01、0.005 和 0.005。

在时间序列预测设定中, 本研究针对每个数据集固定输入长度, 并逐步增加预测窗口的尺寸,

以评估算法在不同预测时段内的表现。通过设定不同的未来预测视野,分别为12、24和48个时间步,检验算法在不同预测长度范围内的预测性能。

3.4 评价指标

为对算法性能进行全面评估,使用两个评价指标衡量基于联邦学习的时间序列预测算法的性能。这两个指标分别是均方误差(mean-square error, E_{MS})和平均绝对误差(mean absolute error, E_{MA})。这两种评价指标广泛应用于评估回归模型的准确性。

E_{MS} 的计算公式如下:

$$E_{MS} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (7)$$

式中: n 代表样本数量,即预测窗口大小; y_i 为真实值; \hat{y}_i 表示模型输出的预测值。当 E_{MS} 越小时,说明模型的预测值和真实值间差距越小。

E_{MA} 在面对较大误差时显示出较高敏感性,在某些情况下容易受到异常值的影响。本研究还采用 E_{MA} 作为评价指标,它与 E_{MS} 不同之处在于 E_{MA} 对异常值的敏

感程度较低。 E_{MA} 的计算公式如下:

$$E_{MA} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|. \quad (8)$$

较高的 E_{MA} 表明模型预测值与真实值之间的差距较大,从而意味着预测模型准确性较差。

3.5 性能分析

表1总结了所有算法在ETTh1数据集上的多变量预测评估结果。表中粗体数字表示除集中式学习方法之外的最佳性能表现。多变量时间序列预测旨在预测时间序列数据中多个相关变量的未来值,与单变量预测相比,多变量预测能更全面考虑多个变量之间的相互影响和关联,更能反映算法的综合性能。为进一步评估算法的预测能力,本研究逐步扩展预测窗口长度。

从表1可以观察到,在不同预测窗口长度下,本研究提出的算法均表现出更高性能。以预测窗口长度为24为例,本算法在 E_{MS} 上相对于FedAvg、FedProx和MOON分别降低4.8%、3.61%和2.76%。本研究设计的算法在预测性能上甚至胜过可能存在数据泄露的集中式学习方法。

表1 数据集ETTh1上时间序列预测结果
Table 1 Time series forecasting results on dataset ETTh1

方法	预测窗口长度 12		预测窗口长度 24		预测窗口长度 48	
	E_{MS}	E_{MA}	E_{MS}	E_{MA}	E_{MS}	E_{MA}
Centralized	0.409 4	0.427 6	0.416 2	0.444 1	0.473 8	0.484 7
FedAvg	0.439 2	0.452 8	0.450 1	0.472 4	0.484 9	0.495 4
FedProx	0.434 6	0.442 4	0.438 2	0.453 1	0.516 9	0.500 8
MOON	0.442 5	0.449 3	0.429 7	0.454 4	0.480 1	0.494 5
Ours	0.409 0	0.439 0	0.402 1	0.446 5	0.473 4	0.486 3

在不同预测窗口长度下,各算法在ETTm1数据集上的均方误差和平均绝对误差值如表2所示。本研究提出的算法在性能方面始终表现出显著优势,始终取得了最低的 E_{MS} 和 E_{MA} ,这反映了其出色的预测性能。与FedAvg相比,本研究提出的算法

采取多重改进措施,包括利用漂移变量对本地模型训练进行约束以防止本地模型漂移,以及对全局模型聚合过程优化。这些改进明显增强了各客户端间的公平性,提高全局模型的泛化能力,从而有效提升算法性能。

表2 数据集ETTm1上时间序列预测结果
Table 2 Time series forecasting results on dataset ETTm1

方法	预测窗口长度 12		预测窗口长度 24		预测窗口长度 48	
	E_{MS}	E_{MA}	E_{MS}	E_{MA}	E_{MS}	E_{MA}
Centralized	0.182 1	0.276 3	0.292 3	0.350 5	0.388 7	0.409 3
FedAvg	0.212 1	0.297 2	0.324 2	0.379 8	0.425 0	0.440 6
FedProx	0.203 2	0.302 3	0.303 5	0.367 0	0.399 8	0.423 0
MOON	0.200 1	0.291 4	0.316 2	0.375 3	0.398 1	0.431 4
Ours	0.195 3	0.289 9	0.287 2	0.357 1	0.371 6	0.416 6

如表3所示,在Weather数据集上使用相同设置,本研究设计的算法在大多数预测窗口长度下都

展现出比其它基线方法更出色的效果,这充分验证算法有效性。

表3 数据集 Weather 上时间序列预测结果
Table 3 Time series forecasting results on dataset Weather

方法	预测窗口长度 12		预测窗口长度 24		预测窗口长度 48	
	E_{MS}	E_{MA}	E_{MS}	E_{MA}	E_{MS}	E_{MA}
Centralized	0.070 5	0.123 7	0.099 6	0.177 9	0.133 7	0.213 2
FedAvg	0.091 4	0.145 0	0.157 5	0.225 0	0.224 7	0.295 9
FedProx	0.077 4	0.140 6	0.134 2	0.200 4	0.203 9	0.291 0
MOON	0.090 6	0.152 2	0.169 4	0.228 6	0.225 3	0.293 5
Ours	0.084 6	0.138 2	0.131 9	0.195 2	0.199 2	0.290 5

3.6 消融试验

本研究设计的算法具备两个关键特点:引入正则化项和采用客户端贡献估计策略。这两方面的改进能显著增强基于联邦学习的时间序列预测算法的性能。为验证这两个模块对算法性能的改进效果,在 ETTh1 上进行 4 组对比试验,其中预测窗口长度设置为 12。试验结果如表 4 所示。

表 4 不同模块的有效性
Table 4 Effectiveness of different modules

正则化项	贡献估计策略	E_{MS}	E_{MA}
×	×	0.439 2	0.452 8
√	×	0.430 9	0.453 5
×	√	0.411 5	0.440 1
√	√	0.409 0	0.439 0

“x”和“√”分别表示在算法中是否使用这两个模块进行试验。从结果可以看出,正则化项和客户端贡献估计策略都会对算法性能产生积极影响,这一结果充分验证本研究提出的模块在性能改进方面的有效性。

3.7 超参数敏感性分析

为评估超参数选择对本研究提出算法的影响,针对式(2)中的超参数 η 进行多组选择试验。在 $[0.001, 1]$ 内选择不同的 η ,并在 3 个公开数据集上进行测试。图 1 展示在 ETTh1 上的测试精度。

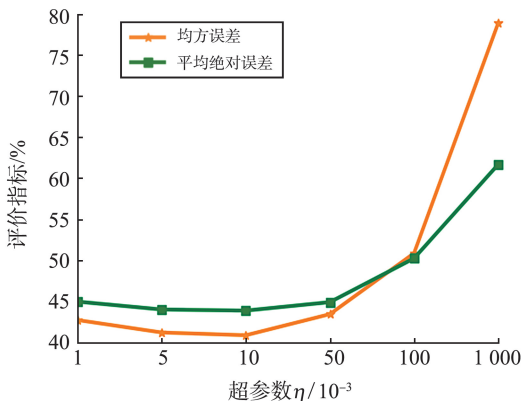


图 1 ETTh1 上超参数 η 对模型性能的影响
Fig.1 The impact of η on E_{MS} and E_{MA} on ETTh1

如图 1 所示,在 ETTh1 数据集上,使用客户端

数量为 5,预测窗口长度为 12 的设置下,当 η 小于 0.05 时,本研究的算法表现出较小敏感性,取得较低的 E_{MS} 和 E_{MA} 。

如图 2 所示,在 ETTm1 数据集上采用相同设置时,本研究提出的算法在 η 小于 0.05 时展现出较为稳定的预测性能。

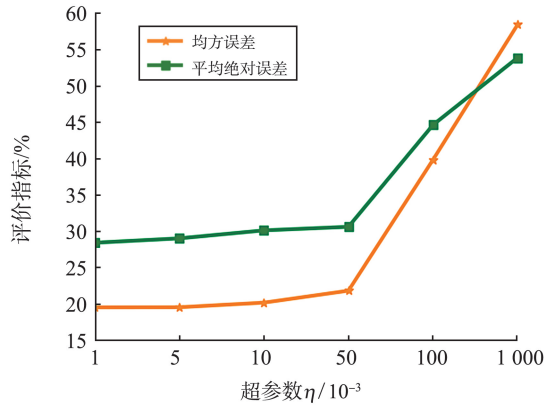


图 2 ETTm1 上超参数 η 对模型性能的影响
Fig.2 The impact of η on E_{MS} and E_{MA} on ETTm1

在 Weather 数据集上,采用 5 个客户端并设置预测窗口长度为 12,算法的性能测试结果如图 3 所示。结果显示,在 η 小于 0.05 时,算法性能表现相对稳定,与前两个数据集的观测一致。这进一步验证本研究提出的算法在超参数选择方面具有较好鲁棒性。

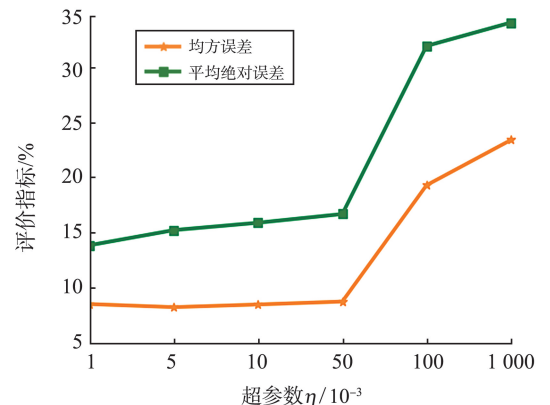


图 3 Weather 上超参数 η 对模型性能的影响
Fig.3 The impact of η on E_{MS} and E_{MA} on Weather

在实际的联邦学习场景中,客户端数量可能会

非常庞大,为深入了解客户端数量对联邦学习算法的影响,本研究在3个数据集上分别进行试验。图4展示在ETTh1数据集上超参数 η 被设置为0.01时,算法在不同客户端数量下的性能结果。可以明显观察到,随着客户端数量的增加,算法性能逐渐下降。

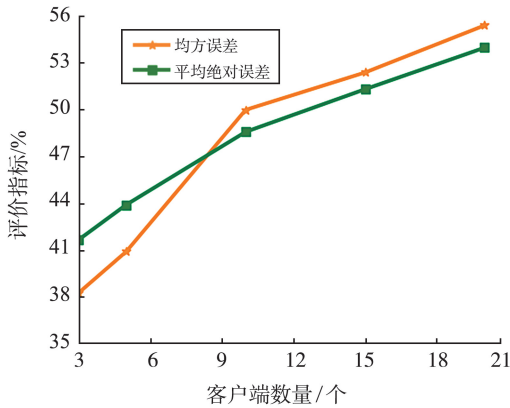


图4 ETTh1上客户端数量对模型性能的影响
Fig.4 The impact of K on performance on ETTh1

图5和图6分别展示在超参数 η 被设置为0.005时,ETTh1和Weather数据集上不同客户端数量对算法性能的影响程度。

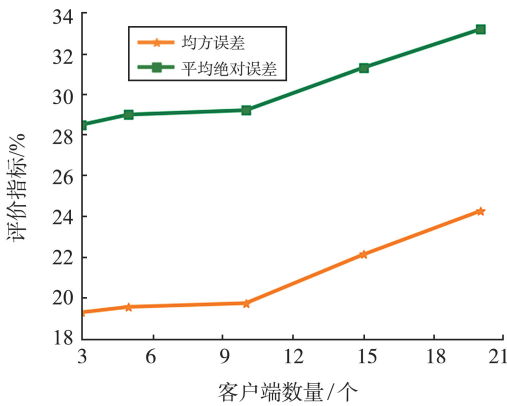


图5 ETTh1上客户端数量对模型性能的影响
Fig.5 The impact of K on performance on ETTh1

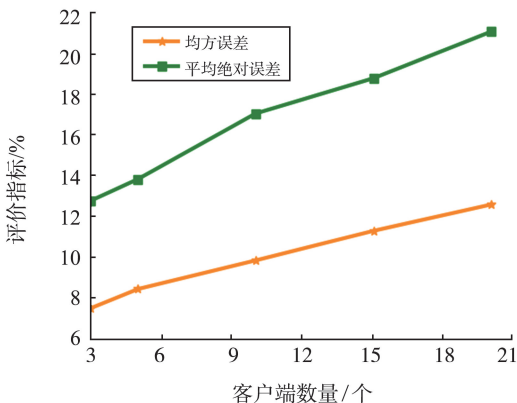


图6 Weather上客户端数量对模型性能的影响
Fig.6 The impact of K on performance on Weather

这一结果表明,客户端数量对联邦学习算法的性能具有重要影响。当客户端数量增加时,会导致模型梯度下降方向变得复杂,影响算法的收敛速度和稳定性。

3.8 训练轮数分析

为更直观地展示本研究提出的算法在性能方面的优势,在ETTh1数据集上,将其与上述基线算法进行比较。在训练过程中,对所有算法设置客户端数量为5,预测窗口长度为12,并确保算法中特定超参数为各自最优设置。如图7所示,展示所有算法随着训练轮次 T 变化的 E_{MS} 。结果表明,本研究提出的算法在预测性能方面优于基线方法,进一步验证本研究提出的算法在性能方面的优势。

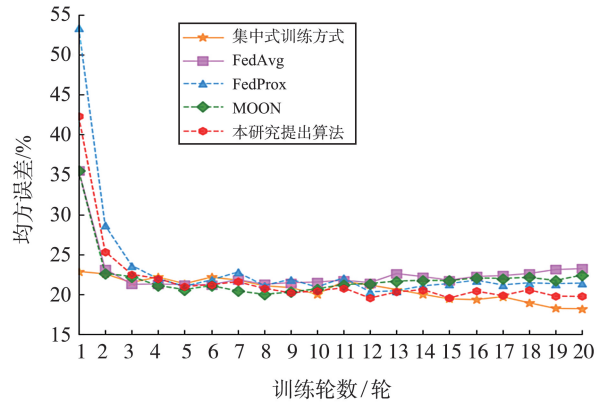


图7 ETTh1上训练轮数与均方误差间的关系
Fig.7 Relationship between T and E_{MS} on ETTh1

图8展示在相同设置下,所有算法中训练轮次 T 与 E_{MA} 之间的关系。更高的预测准确性可以归因于本研究提出的算法具有减轻本地模型漂移和有效优化整体目标的能力。通过优化算法,能够更准确预测时间序列数据,提高模型泛化能力。

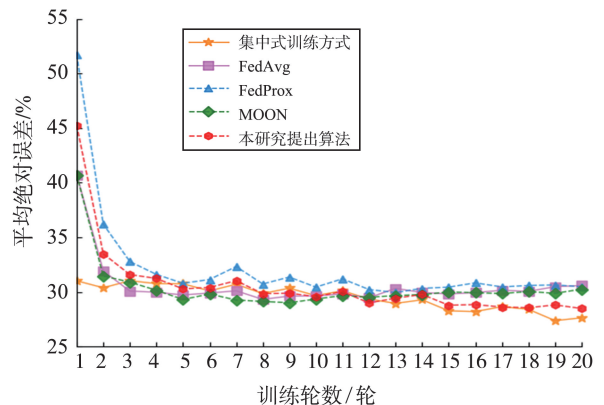


图8 ETTh1上训练轮数与平均绝对误差间的关系
Fig.8 Relationship between T and E_{MA} on ETTh1

4 结论

本研究提出一种全新的联邦时间序列预测算

法。该算法在客户端本地训练过程中引入一个正则化项,使用漂移变量来量化本地模型和全局模型间的差距,用该变量约束本地模型训练,减少本地漂移对全局目标的影响,加快本地模型在训练过程中的收敛速率。在全局模型聚合过程中,提出一种新的客户端贡献估计策略,用于动态调整分配给每个客户端的权重,利用本地更新和全局更新间的余弦相似性来度量客户端的重要性,保证在参数聚合过程中的公平性。通过在客户端本地更新和服务端全局聚合过程中分别引入合适策略,所提出的算法能有效克服数据异质性和不公平性带来的挑战,为分布式时间序列预测任务提供更为高效的解决方案。本研究在多个数据集上进行充分试验,结果表明该算法具有有效性。

参考文献:

- [1] 张美英,何杰. 时间序列预测模型研究综述[J]. 数学的实践与认识, 2011, 41(18): 189-195.
ZHANG Meiyang, HE Jie. A comprehensive review of time series forecasting models[J]. Mathematical Practice and Understanding, 2011, 41(18): 189-195.
- [2] 李英惠,胥超. 基于时间序列模型的售电量预测方法[J]. 山东电力技术, 2014, 41(6): 56-59.
LI Yinghui, XU Chao. Sales Electricity forecasting method based on time series model[J]. Shandong Electric Power Technology, 2014, 41(6): 56-59.
- [3] 杨海民,潘志松,白玮. 时间序列预测方法综述[J]. 计算机科学, 2019, 46(1): 21-28.
YANG Haimin, PAN Zhisong, BAI Wei. A comprehensive review of time series forecasting methods[J]. Computer Science, 2019, 46(1): 21-28.
- [4] 王连成,代桃桃. 数据驱动创新场景引领未来[J]. 山东电力技术, 2018, 45(10): 22-26.
WANG Liancheng, DAI Taotao. Data-driven innovation leading the future of scenarios[J]. Shandong Electric Power Technology, 2018, 45(10): 22-26.
- [5] 吕秋霞,孙亮,车延华,等. 基于深度置信网络的配电网负荷预测[J]. 山东电力技术, 2023, 50(8): 20-26.
Lǚ Qiuxia, SUN Liang, CHE Yanhua, et al. Distribution network load forecasting based on deep belief networks[J]. Shandong Electric Power Technology, 2023, 50(8): 20-26.
- [6] TORRES J F, HADJOUT D, SEBAA A, et al. Deep learning for time series forecasting: a survey[J]. Big Data, 2021, 9(1): 3-21.
- [7] LI T, SAHU A K, TALWALKAR A, et al. Federated learning: challenges, methods, and future directions[J]. IEEE Signal Processing Magazine, 2020, 37(3): 50-60.
- [8] BONAWITZ K, EICHNER H, GRIESKAMP W, et al. Towards federated learning at scale: system design[J]. Machine Learning and Systems, 2019, 1: 374-388.
- [9] 王健宗,孔令炜,黄章成,等. 联邦学习算法综述[J]. 大数据, 2020, 6(6): 64-82.
WANG Jianzong, KONG Lingwei, HUANG Zhangcheng, et al. A comprehensive review of federated learning algorithms[J]. Big Data, 2020, 6(6): 64-82.
- [10] 周传鑫,孙奕,汪德刚,等. 联邦学习研究综述[J]. 网络与信息安全学报, 2021, 7(5): 77-92.
ZHOU Chuanxin, SUN Yi, WANG Degang, et al. A comprehensive review of federated learning[J]. Journal of Network and Information Security, 2021, 7(5): 77-92.
- [11] 陈兵,成翔,张佳乐,等. 联邦学习安全与隐私保护综述[J]. 南京航空航天大学学报, 2020, 52(5): 675-684.
CHEN Bing, CHENG Xiang, ZHANG Jiale, et al. A comprehensive review of security and privacy protection in federated learning[J]. Journal of Nanjing University of Aeronautics and Astronautics, 2020, 52(5): 675-684.
- [12] ZHANG G P. Time series forecasting using a hybrid ARIMA and neural network model[J]. Neurocomputing, 2003, 50: 159-175.
- [13] NING Y, KAZEMI H, TAHMASEBI P. A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and Prophet[J]. Computers & Geosciences, 2022, 164: 105126.
- [14] 林颖,刘萌,白德盟,等. 基于深度学习的电力设备红外可见光图像智能配准方法研究[J]. 山东电力技术, 2022, 49(8): 22-27.
LIN Ying, LIU Meng, BAI Demeng, et al. Research on intelligent registration of infrared and visible light images of power equipment based on deep learning[J]. Shandong Electric Power Technology, 2022, 49(8): 22-27.
- [15] 夏瑜璐. 循环神经网络的发展综述[J]. 电脑知识与技术, 2019, 15(21): 182-184.
XIA Yulu. Development overview of recurrent neural networks[J]. Computer Knowledge and Technology, 2019, 15(21): 182-184.
- [16] 周飞燕,金林鹏,董军. 卷积神经网络研究综述[J]. 计算机学报, 2017, 40(6): 1229-1251.
ZHOU Feiyan, JIN Linpeng, DONG Jun. A comprehensive review of convolutional neural networks[J]. Chinese Journal of Computers, 2017, 40(6): 1229-1251.
- [17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Neural Information Processing Systems, 2017, 30: 5998-6008.

- [18] SALINAS D, FLUNKERT V, GASTHAUS J, et al. DeepAR: probabilistic forecasting with autoregressive recurrent networks[J]. *International Journal of Forecasting*, 2020, 36(3): 1181-1191.
- [19] 张宸嘉, 朱磊, 俞璐. 卷积神经网络中的注意力机制综述[J]. *计算机工程与应用*, 2021, 57(20): 64-72.
ZHANG Chenjia, ZHU Lei, YU Lu. A survey of attention mechanisms in convolutional neural networks [J]. *Computer Engineering and Applications*, 2021, 57(20): 64-72.
- [20] 田永林, 王雨桐, 王建功, 等. 视觉 Transformer 研究的关键问题: 现状及展望[J]. *自动化学报*, 2022, 48(4): 957-979.
TIAN Yonglin, WANG Yutong, WANG Jianguo, et al. Key issues in visual transformer research: current status and prospects[J]. *Acta Automatica Sinica*, 2022, 48(4): 957-979.
- [21] ZHOU H, ZHANG S, PENG J, et al. Informer: Beyond efficient transformer for long sequence time-series forecasting[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. Palo Alto, USA: AAAI Press, 2021: 11106-11115.
- [22] MCMAHAN B, MOORE E, RAMAGE D, et al. Communication efficient learning of deep networks from decentralized data[C]//*Proceedings of the Artificial Intelligence and Statistics*. Florence, Italy: PMLR, 2017: 1273-1282.
- [23] LI T, SAHU A K, ZAHEER M, et al. Federated optimization in heterogeneous networks[C]//*Proceedings of Machine Learning and Systems*. Texas, USA: PMLR, 2020: 429-450.
- [24] LI Q, HE B, SONG D. Model-contrastive federated learning[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Los Alamitos, USA: IEEE, 2021: 10713-10722.
- [25] MENDIETA M, YANG T, WANG P, et al. Local learning matters: rethinking data heterogeneity in federated learning[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA: IEEE, 2022: 8397-8406.
- [26] 王文鑫, 张健毅. 联邦学习公平性研究综述[J]. *北京电子科技学院学报*, 2022, 30(2): 122-134.
WANG Wenxin, ZHANG Jianyi. A comprehensive review of fairness in federated learning[J]. *Journal of Beijing Electronic Science and Technology Institute*, 2022, 30(2): 122-134.
- [27] 田家会, 吕锡香, 邹仁朋, 等. 一种联邦学习中的公平资源分配方案[J]. *计算机研究与发展*, 2022, 59(6): 1240-1254.
TIAN Jiahui, LÜ Xixiang, ZOU Renpeng, et al. A fair resource allocation scheme in federated learning [J]. *Journal of Computer Research and Development*, 2022, 59(6): 1240-1254.
- [28] 张军阳, 王慧丽, 郭阳, 等. 深度学习相关研究综述[J]. *计算机应用研究*, 2018, 35(7): 1921-1936.
ZHANG Junyang, WANG Huili, GUO Yang, et al. A comprehensive review of deep learning research[J]. *Journal of Computer Applications Research*, 2018, 35(7): 1921-1936.
- [29] LUO K, LI X, LAN Y, et al. GradMA: a gradient-memory-based accelerated federated learning with alleviated catastrophic forgetting[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, Canada: IEEE, 2023: 3708-3717.
- [30] WU H, XU J, WANG J, et al. Autoformer: decomposition transformers with auto-correlation for long-term series forecasting[J]. *Neural Information Processing Systems*, 2021, 34: 22419-22430.

(编辑:陈燕)