

电表数据隐私保护下的联邦学习行业电力负荷预测框架

王蓓蓓¹, 朱 竞¹, 王嘉乐², 马璘劼³

(1. 东南大学电气工程学院, 江苏省南京市 210096; 2. 东南大学网络空间安全学院, 江苏省南京市 211189;
3. 国网江苏省电力有限公司营销部, 江苏省南京市 210000)

摘要: 电力市场化改革背景下,供电公司的用电采集系统不对主动配电网运营商开放,同时,未来终端用户更倾向于将用户信息保存在本地以保护自己的隐私,主动配电网运营商需要在无读表权的条件下开展负荷预测等电力业务。为此,选择天气和时间因素作为负荷的关联因素,提出一种面向行业用户读表数据保护的联邦学习负荷预测框架。在此基础上,构建了行业用户数据集,基于长短期时间序列网络(LSTNet)建立负荷预测模型,同时利用FedML框架建立基于联邦学习的分行业负荷预测框架。算例分析表明,所述方法能使同行业的用户在不共享负荷数据的前提下进行联邦训练,在保护用户用电隐私的前提下支撑主动配电网运营商相关业务开展,具有较优的预测性能、较少的模型数量和较短的耗时。

关键词: 长短期时间序列网络; 负荷预测; 联邦学习; FedML框架; 隐私保护

0 引言

随着电力市场化改革的逐步深入,主动配电网运营商进入市场代理用户进行批发市场购电,但用户历史用电曲线数据往往保存在供电公司的用电采集系统中,出于保护用户隐私的需要,供电公司会拒绝向主动配电网运营商共享用电采集数据。此外,未来大量分布式新能源发电和储能等设备接入配电网中,传统配电网正逐步转变为主动配电网。由于攻击者可以借助非侵入式负荷检测算法将用户的总负荷信息分解到具体的用电设备上掌握用户的生活规律^[1]。因此,在未来的主动配电网中,用户会更倾向于将读表数据保存在本地,主动配电网运营商需要在无读表权的条件下开展负荷预测等业务。

传统负荷预测基于时间序列预测的研究而展开,主要包括差分整合移动平均自回归(autoregressive integrated moving average, ARIMA)模型^[2]、向量自回归(vector auto regression, VAR)模型^[3]、随机森林(random forest, RF)^[4]、门控循环单元(gated recurrent unit, GRU)^[5]、极限学习机(extreme learning machine, ELM)^[6]和长短期记忆(long short-term memory, LSTM)网络^[7]等方法。

传统负荷预测根据读表数据基于历史相似日类推等常规预测方法进行预测,如果用户选择将信息保存在本地来保护隐私,则主动配电网运营商无法获取其历史读表数据,负荷预测的精度将大大降低。如何在不掌握用户读表数据的情况下,依然能对用户用电水平进行准确预测就显得尤其重要,而联邦学习(federated learning, FL)为实现考虑隐私保护的负荷预测提供了可能。联邦学习是一种分布式的训练方式,响应用户将负荷数据保留在本地进行训练,只需将训练后的负荷特性参数上传到主动配电网运营商的中央服务器,中央服务器收集多个响应用户的负荷特性模型参数,基于联邦优化算法得到全局模型参数并将其下发给响应用户进行更新,再进行下一轮次的训练,直至获得最终预测模型^[8]。文献[9]考虑了一种基于能源需求学习的预测模型,对电动汽车的充电需求进行预测;文献[10]提出了面向信息安全的电力计量联邦学习框架,帮助电力计量系统中多个部门在保证本部门独立数据安全的前提下开展业务;文献[11]提出了结合深度学习、变分贝叶斯推断和联邦学习的太阳辐射照度联邦概率预测模型;文献[12]提出了一种基于分布式优化思想的配电网用电负荷多层协同预测方法,通过多个深度学习模型的分布式联邦训练,提升所有层级负荷的整体预测准确度;文献[13]研究了基于联邦学习和负荷聚类的短期住宅负荷预测方法,应用K均值聚类算法将住宅用户根据其历史用电模式划分集群,

收稿日期: 2022-03-21; 修回日期: 2022-09-20。

上网日期: 2022-12-15。

教育部人文社科一般项目资助(面向多场景社会经济发展评价的电力数据价值深度挖掘方法研究,21YJAZH083)。

对于每个集群应用联邦学习训练负荷预测模型。上述研究主要针对特定场景和用户展开,泛化能力较弱。

针对上述问题,本文设计了主动配电网面向行业用户读表数据保护的联邦学习负荷预测模型,按工业行业类别分别构建了行业用户数据集和相应的数据预处理方案,基于长短期时间序列网络(long-and short-term time-series network, LSTNet)建立客户本地负荷预测模型^[14],利用联邦平均(federated averaging, FedAvg)算法^[15]进行模型参数聚合,利用FedML框架建立基于联邦学习的分行业电力负荷预测框架,在保护用户数据隐私的前提下得到行业负荷预测模型。算例分析结果表明,本文方案具有较优的预测性能、较少的模型数量和较短的耗时,能在未掌握用户数据的情况下获取到准确的预测模型。

1 基于LSTNet的负荷预测

1.1 负荷的影响因素

负荷曲线可以分解为规律分量、不确定分量和噪声分量^[16]:规律分量为负荷在时间维度上的周期性变化;不确定分量为天气、经济等因素作用导致负荷的非周期性变化;噪声分量为其他无法用物理解释的因素对负荷的影响。常见的天气因素有温度、湿度、降水量和气压等,这些天气因素对短期负荷的变化影响较大;时间因素体现在短期重复模式和长期重复模式的共同作用上,主要包括工作日、节假日和季节3个方面的影响。

1.2 行业用户数据集构建

本文选择天气、负荷和时间作为负荷预测模型输入数据的特征类型:天气包括温度、相对湿度、降水量、风速和气压,用于反映天气因素对负荷的影响;时间数据包括节假日信息、工作日信息、年月日时分及星期信息,用于反映负荷的周期性,其中,节假日信息根据中国2015年和2016年的法定节假日确定,工作日信息包括周中和由于调休产生的工作日。这些构建的行业数据集只需保留在用户本地,无须提供给主动配电网运营商,有效保护了用户的数据隐私。表1描述了数据集的特征选取和特征转换,最终得到共27个特征。

1.3 基于LSTNet的负荷预测模型

本文以LSTNet作为负荷预测模型,LSTNet由卷积层、循环层和循环-跳跃层、全连接层和自回归层3部分组成。LSTNet能够利用卷积层从输入的多变量时间序列中提取多变量的短期模式和多变量之间的短期依赖关系,然后,利用循环层和循环-跳

跃层捕捉多变量之间长期和更长期的依赖关系。模型的输入为多变量时间序列 $X=$

$$\begin{bmatrix} x_0(1) & x_1(1) & \cdots & x_n(1) \\ x_0(2) & x_1(2) & \cdots & x_n(2) \\ \vdots & \vdots & & \vdots \\ x_0(m) & x_1(m) & \cdots & x_n(m) \end{bmatrix}, \text{其中}, n \text{ 表示多变量}$$

时间序列的特征数,此处为负荷、天气和时间共27个特征(见表1); m 表示输入的时间序列长度。

表1 数据集的特征选取和特征转换
Table 1 Feature selection and feature transformation of datasets

特征类型	特征名	数值类型	单位/范围	转换方式	转换后特征
天气	温度	连续值	℃	归一化	temp
	湿度	连续值	%	归一化	humidity
	降雨量	连续值	mm	归一化	precipitation
	气压	连续值	hPa	归一化	barometer
	风速	连续值	m/s	归一化	windspeed
负荷	负荷	连续值	kW	归一化	load
时间	年	离散值	2015/2016	独热编码	year_2015, year_2016
	月	离散值	1~12	sin/cos 循环编码	month_sin, month_cos
	日	离散值	1~30	sin/cos 循环编码	day_sin, day_cos
	时	离散值	1~24	sin/cos 循环编码	hour_sin, hour_cos
	分	离散值	1~60	sin/cos 循环编码	minute_sin, minute_cos
					day_of_week_0, day_of_week_1, day_of_week_2, day_of_week_3, day_of_week_4, day_of_week_5, day_of_week_6
	星期	离散值	0~6	独热编码	
	工作日信息	离散值	是/否	独热编码	workday_0, workday_1
	节假日信息	离散值	是/否	独热编码	holiday_0, holiday_1

记卷积层对应的权重系数为 W_C 、偏置为 b_C ,循环层对应的权重系数为 W_R 、偏置为 b_R ,循环-跳跃层对应的权重系数为 W_S 、偏置为 b_S ,全连接层的偏置为 b_D ,自回归层对应的权重系数为 W_{AR} 、偏置为 b_{AR} 。将LSTNet中各层的权重系数 $W=\{W_C, W_R, W_S, W_{AR}\}$ 和偏置 $b=\{b_C, b_R, b_S, b_D, b_{AR}\}$ 统记为:

$$\omega \in \{W, b\} \quad (1)$$

式中: ω 代表各层神经元的权重系数和偏置, 即 LSTNet 的模型参数。

2 基于联邦学习的分行业负荷预测框架

2.1 联邦学习概述

联邦学习是一种基于分布式数据集的机器学习方法, 最先由 Google 的 McMahan 提出, 联邦学习包括模型训练和模型推理两个过程。模型训练过程中, 多个客户端在本地用各自的数据集进行训练, 由服务器收集模型参数做全局模型的更新, 各客户端基于该更新继续训练, 最终得到可以多方共享的模型。模型推理是将训练好的联邦学习模型应用在新的数据上。文献[17]针对不同客户端的数据在特征空间和样本空间的重叠情况, 将联邦学习分为横向联邦学习、纵向联邦学习和联邦迁移学习。其中, 横向联邦学习适用于客户端在特征空间有重叠而样本空间不同的情况。横向联邦学习的适用条件如式(2)所示。

$$\begin{cases} X_i = X_j \\ I_i \neq I_j \end{cases} \quad \forall D_i, D_j, i \neq j \quad (2)$$

式中: D_i 和 D_j 分别为客户端 i 和客户端 j 的数据集; X_i 和 X_j 分别为客户端 i 和客户端 j 的特征空间; I_i 和 I_j 分别为客户端 i 和客户端 j 的样本空间。

在本文的应用场景中, 相同行业的不同电力终端用户各自拥有自己的数据集, 数据集中的样本不同, 但样本对应的特征类似相同(负荷形状、天气信息、时间信息)。如果未来用户出于隐私保护的角度拒绝共享其读表数据, 主动配电网运营商可以利用本文提出的联邦学习算法框架进行行业用户负荷预测工作, 用户只需在本地利用自己的负荷数据训练出自己的负荷预测模型, 将负荷预测模型参数而非具体的用户读表负荷数据共享给主动配电网运营商, 而主动配电网运营商利用 FedAvg 算法对这些模型参数进行聚合, 利用 FedML 框架建立基于联邦学习的分行业电力负荷预测框架, 在保护用户数据隐私的前提下得到该行业的负荷预测模型。

2.2 基于 FedAvg 算法的联邦训练流程

联邦学习基于 FedAvg 算法获得全局模型。FedAvg 算法是 McMahan 提出的横向联邦优化算法, 通过在多个用户侧的客户端并行迭代运行梯度下降算法, 并在每一轮交互中由中央服务器收集参数进行聚合后发回给客户端继续训练。

参与联邦学习的用户为响应用户, 模型参数为 LSTNet 各层神经元的权重系数或偏置。主动配电网运营商/供电公司设置中央服务器用于聚合响应

用户上传的本地模型参数, 响应用户使用本地数据集基于 LSTNet 负荷预测模型进行本地训练, 彼此之间不共享数据集。本地训练达到设置的训练批次后, 响应用户将本地模型参数上传到中央服务器。在收集完所有参与联邦训练的响应用户上传的本地模型参数后, 中央服务器基于 FedAvg 算法生成全局模型参数后发回给各响应用户。响应用户用全局模型参数更新自己的本地模型参数后继续训练, 重复这一过程直至达到设置的交互轮次。具体的流程如图 1 所示。

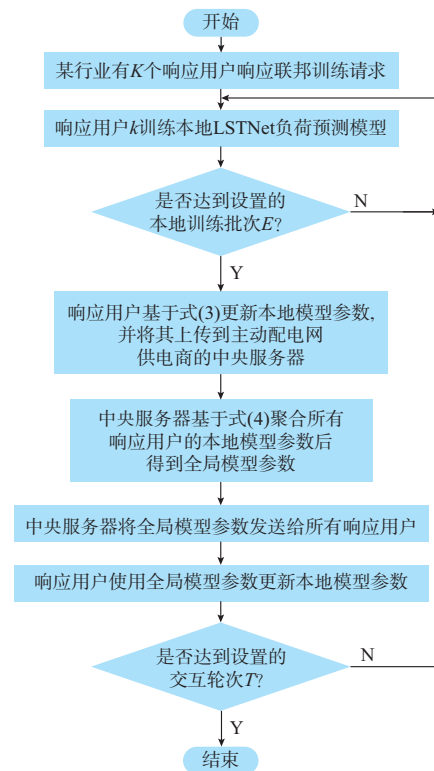


图 1 基于 FedAvg 算法的联邦训练流程图
Fig. 1 Flow chart of federated training based on FedAvg algorithm

首先, 位于主动配电网运营商的中央服务器将初始化后的模型参数 ω_0 发送给所有响应用户。响应用户接收到模型参数后, 在本地基于梯度下降算法训练 LSTNet 模型。在第 t 轮交互中, 响应用户 k 首先进行本地训练, 训练完成后的模型参数 $\omega_{k,t}$ 如式(3)所示。

$$\omega_{k,t} = \omega_{k,t-1} - \eta \nabla f(\omega_{k,t-1}) \quad (3)$$

式中: $\omega_{k,t}$ 为响应用户 k 在第 t 轮交互中本地训练完成后 LSTNet 各层神经元的权重系数或偏置; η 为学习率; ∇f 表示梯度下降。

然后, 各响应用户将式(3)中训练完成的模型参数 $\omega_{k,t}$ 发送给位于主动配电网运营商/供电公司的

中央服务器。中央服务器在收集到所有响应用户更新后的模型参数 $\omega_{k,t}$ 后,对参数按式(4)做聚合处理生成全局参数。

$$\omega_t = \sum_{k=1}^K \frac{n_k}{n} \omega_{k,t} \quad (4)$$

式中: n_k 为响应用户 k 的样本数量; n 为所有响应用户的样本总数; K 为响应用户总数。中央服务器将全局参数 ω_t 发送给所有响应用户,响应用户基于全局参数更新本地参数后继续训练直至完成全部交互轮次,即

$$\omega_{k,t} = \omega_t \quad (5)$$

2.3 基于联邦学习的分行业负荷预测框架

基于联邦学习的分行业负荷预测框架主体包括主动配电网运营商/供电公司、中央服务器、用户以及售电公司。整个流程分为以下7个步骤:

步骤1:主动配电网运营商/供电公司向某目标行业对应的用户发出训练请求。

步骤2:考虑到用户需要在本地完成训练,用户可以决定是否参与到联邦训练中。参与的用户为响应用户,响应用户在本地预处理自己的数据集并将数据集输入LSTNet负荷预测模型中进行本地训练,完成一轮训练后将LSTNet各层的神经元参数传递给中央服务器。

步骤3:中央服务器基于FedAvg算法聚合响应用户传递的本地模型参数生成全局模型参数,并将全局模型参数传递给响应用户,在多次交互后中央服务器获得行业全局模型。

步骤4:中央服务器将行业全局模型传递给主动配电网运营商/供电公司。

步骤5:主动配电网运营商/供电公司根据响应用户对行业全局模型的贡献程度分发相应的奖励。

步骤6:售电公司根据所需的行业,向主动配电网运营商/供电公司提出模型需求申请。

步骤7:主动配电网运营商/供电公司返回所需行业全局模型并获取收益。

3 算例分析

本章通过算例验证所提方案的有效性。采用的实验环境描述如下:CPU为3.70 GHz Intel Core i7-8700K, GPU为NVIDIA TITAN XP 12 GB,内存为32 GB,操作系统选用Ubuntu 18.04 LTS, Pytorch版本为1.8.1, CUDA版本为11.2, Python版本为3.7,联邦学习基于FedML库实现,联邦优化算法为FedAvg算法,采用单机部署模式。用于负荷预测的LSTNet模型参数设置如表2所示。

表2 LSTNet模型参数
Table 2 LSTNet model parameters

参数名称	参数值
卷积层隐藏单元数	128个
卷积层卷积核宽度	6
卷积层中卷积核长度	27
循环层隐藏单元数	128个
循环-跳跃层隐藏单元数	128个
循环-跳跃层跳过隐藏单元数	48个
学习率	0.000 1
丢失率	0.2
一批样本数	64
训练批次	200次

3.1 数据集描述

算例选择的数据集为中国江苏省南京市2015年1月1日—2016年12月31日共731 d按行业分类的多元特征数据集,时间颗粒度设为每天48个点。在构建的数据集中分别从医药制造业、食品制造业、橡胶和塑料制品业3个行业中随机选取3个响应用户。每个响应用户的数据集按照8:1:1的比例(分别对应585、73、73 d)划分为训练集、验证集和测试集。

3.2 评价指标

为衡量模型的预测性能,需要利用评价指标来计算模型的预测误差。常见的评价指标有平均绝对值误差(mean absolute error, MAE)、均方根误差(root mean squared error, RMSE)、平均绝对百分比误差(mean absolute percentage error, MAPE)和对称平均绝对百分比误差(symetric mean absolute percentage error, SMAPE)。本文选取RMSE和SMAPE(其值分别为 λ_{RMSE} 和 λ_{SMAPE})作为短期负荷预测结果的评价指标,表达式分别如式(6)和式(7)所示。

$$\lambda_{\text{RMSE}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (6)$$

$$\lambda_{\text{SMAPE}} = \frac{200}{N} \sum_{i=1}^N \frac{|\hat{y}_i - y_i|}{|\hat{y}_i| + |y_i|} \quad (7)$$

式中: N 为样本总数; \hat{y}_i 和 y_i 分别为第 i 时刻的预测值和实际值。RMSE利用平方项放大了较大误差和较小误差之间的差距,使得RMSE对预测偏差较大的数据更为敏感, RMSE越小表示模型预测准确性越好。SMAPE是MAPE的修正版,解决了在实际值为0时MAPE无法计算以及MAPE对负误差惩罚比正误差更大的问题。SMAPE的取值范围为 $[0, 200]$, SMAPE越小表示预测性能越好。

3.3 联邦学习调优结果分析

本节根据表3设置联邦学习训练场景。其中,交互轮次是指每个行业的中央服务器向各响应用户更新全局参数的次数,而本地训练批次是指各响应用户本地迭代的次数。基于行业用电特性的负荷预测以行业为单位训练模型,本节设置有3个行业,分别为医药制造业、食品制造业、橡胶和塑料制品业,每个行业使用3个响应用户的数据集,基于FedAvg算法捕捉单一行业的行业共性用电特性,得到最终的行业负荷预测模型。

表3 联邦学习训练场景设置
Table 3 Training scenario setting of federated learning

场景	交互轮次/次	本地训练批次/次	是否分行业
1	10	15	是
2	10	20	是
3	15	10	是
4	20	10	是
5	20	10	否

由表4和表5可知,场景4的预测性能最佳。对比场景1~4可知,在总训练次数相同的情况下,交互轮次比本地训练批次更加影响负荷预测的性能。对比场景4和场景5可知,分行业的场景由于捕捉到行业的用电特性,预测性能比不分行业的场景更优。

表4 不同场景下联邦学习预测准确率 RMSE 对比
Table 4 Comparison of RMSE for prediction accuracy of federated learning in different scenarios

场景	RMSE/kW		
	医药制造业	食品制造业	橡胶和塑料制品业
1	131.69	79.56	162.57
2	130.38	80.79	166.41
3	129.87	80.01	167.35
4	127.10	76.53	160.75
5	131.77	84.38	178.32

表5 不同场景下联邦学习预测准确率 SMAPE 对比
Table 5 Comparison of SMAPE for prediction accuracy of federated learning in different scenarios

场景	SMAPE/p.u.		
	医药制造业	食品制造业	橡胶和塑料制品业
1	3.59	5.67	4.52
2	3.55	5.84	4.50
3	3.54	5.71	4.64
4	3.49	5.44	4.38
5	3.59	5.98	4.89

将场景4应用于更多行业以测试本文方法在不同行业下的泛化性和适应性。不同行业下的预测准确率 SMAPE 如图2所示。

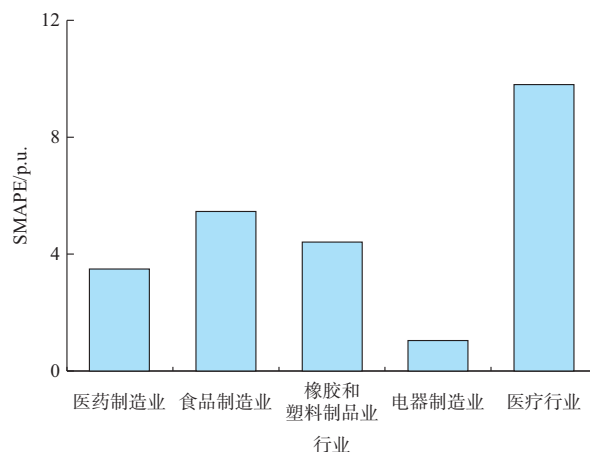


图2 不同行业预测准确率 SMAPE 对比
Fig. 2 Comparison of SMAPE for prediction accuracy in different industries

从图2可以看出,本文所提基于联邦学习的分行业负荷预测框架在不同行业下的预测准确率 SMAPE 值均小于 10 p.u.,而 SMAPE 的理论最大值为 200 p.u.,说明本文方法具有普适性和通用性,可以在不同行业中应用。

3.4 本文方案和不考虑隐私保护方案对比分析

3.4.1 分行业预测方案

在不考虑隐私保护的分行业预测方案中,所有用户都需上传原始读表数据。智能电表通过网关将用户的本地读表数据上传到主动配电网运营商/供电公司的服务器中,分行业进行累加,主动配电网运营商/供电公司基于所有用户的累加数据建立行业负荷预测模型。本节以表3中的场景4为例,与不考虑隐私保护的分行业预测方案进行对比分析。为方便表述,下文以方案1代指不考虑隐私保护的分行业预测方案。

表6和表7为本文所提联邦学习方案与方案1的预测准确率对比。可知,在预测行业用户累加后的负荷的情况下,本文方案和方案1的表现相近,说明本文所提出的考虑隐私保护的分行业电力负荷联邦学习预测模型和传统不考虑隐私保护的分行业独立预测方案性能相近,但本文方案无须向主动配电网运营商/供电公司上传原始读表数据,有效保护了用户的隐私。

图3—图5分别为方案1和本文方案中某两天的预测值对比。

表6 本文方案和方案1预测准确率RMSE对比
Table 6 Comparison of RMSE for prediction accuracy between proposed scheme and scheme 1

方案	RMSE/kW		
	医药制造业	食品制造业	橡胶和塑料制品业
方案1	126.58	79.55	158.81
本文方案(场景4)	127.10	76.53	160.75

表7 本文方案和方案1预测准确率SMAPE对比
Table 7 Comparison of SMAPE for prediction accuracy between proposed scheme and scheme 1

方案	SMAPE/p.u.		
	医药制造业	食品制造业	橡胶和塑料制品业
方案1	3.44	5.67	4.37
本文方案(场景4)	3.49	5.44	4.38

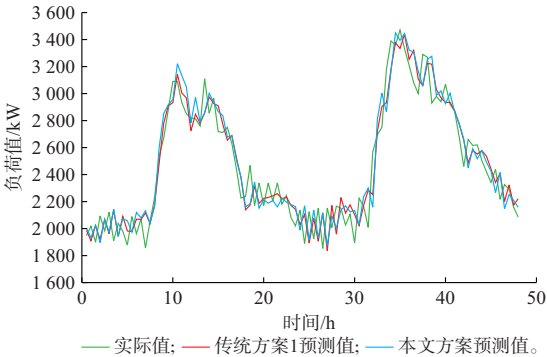


图3 医药制造业方案1和本文方案两天的负荷预测值对比

Fig. 3 Comparison of two-day load forecasting values between scheme 1 and proposed scheme in pharmaceutical manufacturing industry

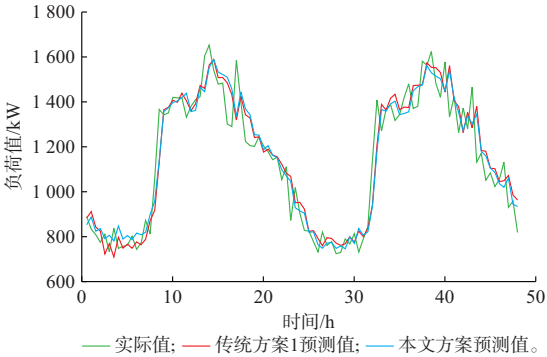


图4 食品制造业方案1和本文方案两天的负荷预测值对比
Fig. 4 Comparison of two-day load forecasting values between scheme 1 and proposed scheme in food manufacturing industry

3.4.2 分用户独立预测方案

在不考虑隐私保护的分用户独立预测方案中,用户需上传原始读表数据,主动配电网运营商/供电

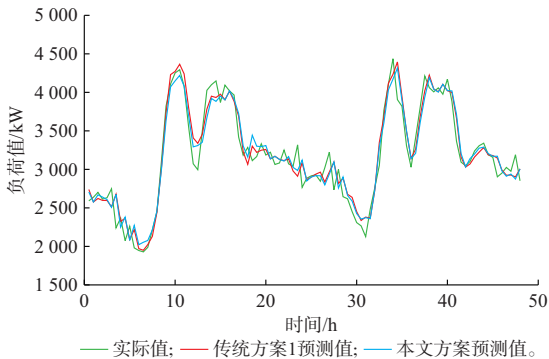


图5 橡胶和塑料制品业方案1和本文方案两天的负荷预测值对比

Fig. 5 Comparison of two-day load forecasting values between scheme 1 and proposed scheme in rubber and plastic products industry

公司针对每个用户的读表数据建立不同的预测模型。本节以表3中的场景4为例,与不考虑隐私保护的分用户独立预测方案进行对比分析。为方便表述,下文以方案2指代不考虑隐私保护的分用户独立预测方案。

表8和表9比较了本文方案与方案2的预测性能。可知,本文方案准确性略逊于方案2,但本文方案考虑了隐私保护,同时仅需建立3个行业预测模型,在计算速度和内存占用方面具有优势。

表8 本文方案和方案2预测准确率RMSE对比
Table 8 Comparison of RMSE for prediction accuracy between proposed scheme and scheme 2

方案	RMSE/kW		
	医药制造业	食品制造业	橡胶和塑料制品业
方案2	128.67	75.74	157.26
本文方案(场景4)	127.10	76.53	160.75

表9 本文方案和方案2预测准确率SMAPE对比
Table 9 Comparison of SMAPE for prediction accuracy between proposed scheme and scheme 2

方案	SMAPE/p.u.		
	医药制造业	食品制造业	橡胶和塑料制品业
方案2	3.55	5.49	4.31
本文方案(场景4)	3.49	5.44	4.38

3.4.3 综合对比分析

表10综合比较了本文方案和不考虑隐私保护的方案。

在预测性能上,本文方案与方案1相近,略逊于方案2。在模型数量上,本文方案与方案1相同,均为行业用户数量,而与方案2相比(方案2的模型数量与用户数量相同),本文方案的模型数量相对更少,在参与训练的用户数增加的情况下优势更加明显。在传输数据量上,传统方案只需传输用户的读

表数据,传输数据量为用户数量(9)×特征数量(27)×天数(731)×一天内的负荷点数(48)×浮点数大小(8),而本文方案中每个响应用户需要在每轮交互时向中央服务器传送本地模型参数,传输数据量为用户数量(9)×模型参数量×交互轮次(20)×浮点数大小(8),模型参数量非常大,仅卷积层、循环层和循环-跳跃层的参数就有 $128 \times 128 \times 48$ 个。本文方案由于考虑了隐私保护,故产生了大量的通信数据,传输成本与传统方案相比较。在耗时上,单次训练用时平均为93 min,方案1需要针对每个行业的数据集建立模型,故耗时为行业数量(3)×单次训练用时(93 min);方案2需要针对每个用户的数据集建立模型,故耗时为用户数量(9)×单次训练用时(93 min);而本文方案训练方式为并行方式,虽然增加了2 min的交互用时,但交互用时相比于训练用时占比较小,耗时为交互用时(2 min)+单次训练用时(93 min)。综上所述,本文方案虽然增大了传输数据量和传输成本,但保证了读表数据的隐私,保持了较优的预测性能、较少的模型数量和较短的耗时。

表10 本文方案和传统方案的综合对比
Table 10 Comprehensive comparison between proposed scheme and traditional schemes

方案	是否考虑隐私保护	SMAPE/p.u.			模型数量/个	传输数据量/MB	耗时/min
		医药制造业	食品制造业	橡胶和塑料制品业			
方案1	否	3.44	5.67	4.37	3	65.05	279
方案2	否	3.55	5.49	4.31	9	65.05	837
本文方案	是	3.49	5.44	4.38	3	1 080	95

4 结语

本文从面向读表数据保护的角度出发,提出了基于联邦学习的分行业电力负荷预测框架。首先,选择天气、经济和时间因素作为负荷主要的影响因素构建数据集;然后,基于LSTNet建立负荷预测模型;最后,利用FedAvg算法和FedML框架建立了基于联邦学习的分行业负荷预测框架。算例分析表明,主动配电网运营商能在未掌握用户数据的情况下得到预测模型,虽然增大了传输数据量,但具有较优的预测性能、较少的模型数量和较短的耗时。

联邦学习对通信方面要求较高,在所需的全局模型规模较大的情况下,网络带宽限制和工作节点数量可能会加剧联邦学习的通信瓶颈。后续可以对天气特征进行特征筛选确定最优天气特征以减少无关气象特征干扰,减少模型传输数据量和提高预测

准确率,增强预测模型泛化能力。同时,进一步考虑如何量化和存证用户对模型的贡献程度,基于响应用户对于预测模型的贡献程度分发奖励,采用经济手段激励更多用户主动参与联邦学习,使得用户得到了经济收益,而主动配电网运营商掌握了更为准确的行业负荷预测模型,形成良性循环。

参考文献

- [1] 肖江, AUGER F, 荆朝霞, 等. 基于贝叶斯信息准则的非侵入式负荷事件检测算法[J]. 电力系统保护与控制, 2018, 46(22): 8-14.
XIAO Jiang, AUGER F, JING Zhaoxia, et al. Non-intrusive load event detection algorithm based on Bayesian information criterion[J]. Power System Protection and Control, 2018, 46(22): 8-14.
- [2] 李妮, 江岳春, 黄珊, 等. 基于累积式自回归动平均传递函数模型的短期负荷预测[J]. 电网技术, 2009, 33(8): 93-97.
LI Ni, JIANG Yuechun, HUANG Shan, et al. Short-term load forecasting based on ARIMA transfer function model[J]. Power System Technology, 2009, 33(8): 93-97.
- [3] TAVEEAPIRADEECHAROEN P, AUNSRI N. A large dimensional VAR model with time-varying parameters for daily forex forecasting[J]. Wireless Personal Communications, 2020, 115(4): 3051-3072.
- [4] ZHU F Y, WU G Q. Load forecasting of the power system: an investigation based on the method of random forest regression[J]. Energy Engineering, 2021, 118(6): 1703-1712.
- [5] FAN C, WANG J Y, GANG W J, et al. Assessment of deep recurrent neural network-based strategies for short-term building energy predictions[J]. Applied Energy, 2019, 236: 700-710.
- [6] SONG G, DAI Q. A novel double deep ELMs ensemble system for time series forecasting[J]. Knowledge-Based Systems, 2017, 134: 31-49.
- [7] JIN X B, YU X H, WANG X Y, et al. Prediction for time series with CNN and LSTM[C]// Proceedings of the 11th International Conference on Modelling, Identification and Control (ICMIC2019), July 13-15, 2019, Tianjin, China: 631-641.
- [8] 杨强. AI与数据隐私保护: 联邦学习的破解之道[J]. 信息安全研究, 2019, 5(11): 961-965.
YANG Qiang. AI and data privacy protection: the way to federated learning[J]. Journal of Information Security Research, 2019, 5(11): 961-965.
- [9] SAPUTRA Y M, HOANG D T, NGUYEN D N, et al. Energy demand prediction with federated learning for electric vehicle networks[C]// 2019 IEEE Global Communications Conference (GLOBECOM), December 9-13, 2019, Waikoloa, USA: 1-6.
- [10] 郑楷洪, 肖勇, 王鑫, 等. 一个面向电力计量系统的联邦学习框架[J]. 中国电机工程学报, 2020, 40(增刊1): 122-133.
ZHENG Kaihong, XIAO Yong, WANG Xin, et al. A federated learning framework for power grid metering system[J]. Proceedings of the CSEE, 2020, 40(Supplement 1): 122-133.

- [11] ZHANG X N, FANG F, WANG J Q. Probabilistic solar irradiation forecasting based on variational Bayesian inference with secure federated learning [J]. IEEE Transactions on Industrial Informatics, 2021, 17(11): 7849-7859.
- [12] 谭嘉,李知艺,杨欢,等.基于分布式优化思想的配电网用电负荷多层协同预测方法[J].上海交通大学学报,2021,55(12): 1544-1553.
TAN Jia, LI Zhiyi, YANG Huan, et al. A multi-level collaborative load forecasting method for distribution networks based on distributed optimization [J]. Journal of Shanghai Jiao Tong University, 2021, 55(12): 1544-1553.
- [13] HE Y, LUO F J, RANZI G, et al. Short-term residential load forecasting based on federated learning and load clustering[C]// 2021 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), October 25-28, 2021, Aachen, Germany: 77-82.
- [14] LAI G K, CHANG W C, YANG Y M, et al. Modeling long- and short-term temporal patterns with deep neural networks [C]// The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, July 8-12, 2018, Ann Arbor, USA: 95-104.
- [15] MCMAHAN B, MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data [C]// 20th International Conference on Artificial Intelligence and Statistics, April 20-22, 2017, Fort Lauderdale, USA: 1273-1282.
- [16] SHI H, XU M H, LI R. Deep learning for household load forecasting—a novel pooling deep RNN[J]. IEEE Transactions on Smart Grid, 2018, 9(5): 5271-5280.
- [17] YANG Q, LIU Y, CHENG Y, et al. Federated learning [J]. Synthesis Lectures on Artificial Intelligence and Machine Learning, 2019, 13(3): 1-207.

王蓓蓓(1979—),女,通信作者,博士,副教授,主要研究方向:电力市场、需求侧管理等。E-mail:wangbeibei@seu.edu.cn

朱 竞(1999—),男,硕士研究生,主要研究方向:电力市场、需求侧管理等。E-mail:zhujing@seu.edu.cn

王嘉乐(1995—),男,硕士研究生,主要研究方向:区块链、隐私保护等。E-mail:gavin0701@gmail.com

(编辑 章黎)

Federated-learning Based Industry Load Forecasting Framework Under Privacy Protection of Meter Data

WANG Beibei¹, ZHU Jing¹, WANG Jiale², MA Jinjie³

(1. School of Electrical Engineering, Southeast University, Nanjing 210096, China;

2. School of Cyber Science and Engineering, Southeast University, Nanjing 211189, China;

3. Marketing Department of State Grid Jiangsu Electric Power Co., Ltd., Nanjing 210000, China)

Abstract: On the background of electricity market reform, the electricity consumption acquisition system of power supply companies is not open to the active distribution network operators. Meanwhile, end users prefer to keep user information locally to protect their privacy in the future, and active distribution network operators need to carry out power business such as load forecasting without meter reading right. Therefore, by selecting the weather and time factors as the correlation factors of load, a federated-learning based load forecasting framework for the protection of meter reading data of industrial users is proposed. On this basis, the industrial user data set is constructed; the load forecasting model is established based on the long- and short-term time-series network (LSTNet); and the sub-industry load forecasting framework based on the federated learning is established by using FedML framework. The case analysis shows that the proposed method can enable users in the same industry to conduct federated training without sharing load data, and support active distribution network operators to carry out relevant business on the premise of protecting users' electricity privacy. It has better prediction performance, fewer models and shorter time consumption.

This work is supported by General Humanities and Social Sciences Research Project of the Ministry of Education (No. 21YJAZH083).

Key words: long- and short-term time-series network (LSTNet); load forecasting; federated learning; FedML framework; privacy protection

