

基于深度神经网络与联邦学习的污染物浓度预测二次建模

钱栋炜, 崔阳光, 魏同权

引用本文

钱栋炜, 崔阳光, 魏同权. 基于深度神经网络与联邦学习的污染物浓度预测二次建模[J]. 计算机科学, 2022, 49(11A): 211200084-5.

QIAN Dong-wei, CUI Yang-guang, WEI Tong-quan. [Secondary Modeling of Pollutant Concentration Prediction Based on Deep Neural Networks with Federal Learning](#) [J]. Computer Science, 2022, 49(11A): 211200084-5.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[深度神经网络的对抗攻击及防御方法综述](#)

Survey of Adversarial Attacks and Defense Methods for Deep Neural Networks
计算机科学, 2022, 49(11A): 210900163-11. <https://doi.org/10.11896/jsjcx.210900163>

[群体智能中的协作与对抗](#)

Cooperation and Confrontation in Crowd Intelligence
计算机科学, 2022, 49(11A): 210900249-7. <https://doi.org/10.11896/jsjcx.210900249>

[变分推断域适配驱动的城市街景语义分割](#)

Variational Domain Adaptation Driven Semantic Segmentation of Urban Scenes
计算机科学, 2022, 49(11): 126-133. <https://doi.org/10.11896/jsjcx.220500193>

[基于安全多方计算和差分隐私的联邦学习方案](#)

Federated Learning Scheme Based on Secure Multi-party Computation and Differential Privacy
计算机科学, 2022, 49(9): 297-305. <https://doi.org/10.11896/jsjcx.210800108>

[基于分层抽样优化的面向异构客户端的联邦学习](#)

Federated Learning Based on Stratified Sampling Optimization for Heterogeneous Clients
计算机科学, 2022, 49(9): 183-193. <https://doi.org/10.11896/jsjcx.220500263>

基于深度神经网络与联邦学习的污染物浓度预测二次建模

钱栋炜 崔阳光 魏同权

华东师范大学计算机科学与技术学院 上海 200062

(469654952@qq.com)

摘要 进入新世纪,伴随着我国经济的高速发展,我国很多地区空气污染情况相对严重,同时政府对于空气污染情况的关注度与治理力度也越来越高。当前对于我国空气质量影响最大的是 O_3 , SO_2 , NO_2 , CO , PM_{10} , $PM_{2.5}$ 这 6 种污染物,因而对这 6 种污染物浓度进行预测预报,及时作出相应管控调整就成为了保障居民健康、建设美丽中国的迫切需求。目前污染物预测的主流方案是 WRF-CMAQ 预测系统,该系统基于污染物物化反应与气象模拟两部分构成。但因为当前对于如臭氧在内的污染物的生成机理等研究还有待深入,WRF-CMAQ 模型的预测存在较大误差。因此采用了深度神经网络对污染物浓度进行二次建模的方式,来减少预测误差。同时,采用联邦学习方法,对于多个监测站使用联邦学习进行数据训练,提升模型泛化能力。实验结果表明,相比于一次 WRF-CMAQ 的一次预测结果,深度神经网络的方案在均方误差值上最多缩小到了 3.93%。同时,采用联邦学习的方案相比于单个监测站点在广泛测试中最多提升了 68.89% 的性能。

关键词 联邦学习;深度神经网络;污染物浓度预测;WRF-CMAQ

中图分类号 TP183

Secondary Modeling of Pollutant Concentration Prediction Based on Deep Neural Networks with Federal Learning

QIAN Dong-wei, CUI Yang-guang and WEI Tong-quan

College of Computer Science and Technology, East China Normal University, Shanghai 200062, China

Abstract In the new century, along with the rapid development of Chinese economy, air pollution in many areas of China is relatively serious, while the government is paying more and more attention to air pollution, and its efforts to control air pollution are increasing. Currently, six pollutants that have the greatest impact on China's air quality are O_3 , SO_2 , NO_2 , CO , PM_{10} , $PM_{2.5}$. Therefore, predicting and forecasting the concentrations of the six pollutants and making corresponding control adjustments in time have become the urgent needs to protect the health of residents and build a beautiful China. At present, the mainstream solution for pollutant prediction is WRF-CMAQ prediction system, which is based on two parts, physical and chemical reaction of pollutants and meteorological simulation. However, due to the current research on the generation mechanism of pollutants such as ozone is still on the way, the prediction of WRF-CMAQ model has large errors. Therefore, this paper adopts a deep neural network for secondary modeling of pollutant concentrations to reduce the prediction error. At the same time, this paper adopts the federal learning method, and uses federal learning for data training for multiple monitoring stations to improve the model generalization ability. Experiment results show that the deep neural network scheme reduces the mean square error value to at most 3.93% compared to the primary prediction results of one WRF-CMAQ. Moreover, the scheme with federal learning improves the performance by up to 68.89% compared to a single monitoring site in extensive tests.

Keywords Federated learning, Deep neural networks, Pollutant concentration prediction, WRF-CMAQ

1 引言

随着中国经济转型升级不断深入,人们对于空气质量的关注程度越来越高。精准预报空气质量情况、结合地区特点综合治理、合理限制污染物排放都离不开精准的预测模型。当前主流的空气质量预报系统的核心是 WRF-CMAQ^[1-5] (Weather Research and Forecasting-Community Multiscale Air Quality) 系统,这也是现在最广泛使用的大规模预报雾霾天气、规划综合治理的核心技术。但是受限于模拟气象场的局限性、排放清单的不确定性、污染物形成机理的复杂性,WRF-CMAQ 系统的预测性能并不理想^[2]。本文希望在该

系统一次预测的基础上,结合所有已知测量值,利用人工智能、深度神经网络、联邦学习,进行二次预测(也可以看作是对一次预测的修正),以期取得比一次预测更准确的结果。

WRF-CMAQ 模型是当前预测空气污染情况的核心技术,该模型包括两个部分:WRF 预测数值天气预报来为 CMAQ 提供所需的气象场数据^[3];CMAQ 是一种化学反应与传输模拟系统,即根据 WRF 提供的气象数据和已知的污染物物理、化学变化机理,推算污染物变化情况的模拟系统。但在真实场景下,WRF 模拟的气象场准确率并不高,同时不同地区的污染物直接排放种类、排放量、排放时间存在一定的差异,再加上 O_3 在内的污染物生成的化学反应机理目前尚未

基金项目:国家自然科学基金(62272169)

This work was supported by the National Natural Science Foundation of China(62272169).

通信作者:魏同权(tqwei@cs.ecnu.edu.cn)

完全探明^[5],因此 WRF-CMAQ 模型预测的性能并不理想。而改进 WRF-CMAQ 模型需要气象学理论更新、物理化学等基础学科的推进^[6](如污染物反应机理)以及对当地实际污染物特点的考察,这些都需要很长的时间积累和大量的实地考察作铺垫,短时间内很难有突破性进展。

人工智能(AI)技术的不断完善给予了空气预测研究新的希望。目前深度神经网络(DNN)^[7]无论从理论上还是应用上已经日趋成熟,类似算法只要有足够数据就能进行训练,最终获得一个具有相当精度的预测模型。但是具体到污染物预测领域,因为原始的数据较为单薄,仅使用原始污染物浓度来进行预测的效果显然不理想。而 WRF-CMAQ 模型虽然在污染物浓度上有不小的偏差,但是它提供了初步模拟气象场、污染物物化反应的参考,这对神经网络模型具有极大的参考价值^[8]。因而,在 WRF-CMAQ 模型的基础上,结合实测数据进行二次建模具有极大的研究和应用价值。同时,拥有足够的污染物实测数据的监测点是有限的,而很多预测需求的预报站大多没有足够的实测数据进行训练,因而本文希望训练的模型具有良好的泛化能力^[9],最终模型能够迁移到更多的监测站上。

为了解决 WRF-CMAQ 模型准确性不足的问题,本文采用深度神经网络和联邦学习算法,对 WRF-CMAQ 模型的初步预测结果进行二次建模。本文主要的贡献点如下:

(1)针对单个气象观测点,引入了深度神经网络模型,在 WRF-CMAQ 模型预测的基础上进行污染物数据的二次建模。

(2)针对多个气象观测点,引入了联邦学习算法,通过联邦学习来提升多个相关观测点预测模型的泛化能力。

(3)广泛的实验表明,本文采用的二次建模方法及联邦学习算法在实验测试中取得了最佳的效果。相比于一次 WRF-CMAQ 的一次预测结果,深度神经网络的方案均方误差最多缩小到了 3.93%。同时,采用联邦学习的方案相比于单个监测站点最多提升了 68.89% 的性能。

本文第 2 节介绍相关工作;第 3 节介绍本文提出的算法;第 4 节给出实验结果;最后总结全文。

2 相关工作

2.1 WRF-CMAQ 模型

WRF-CMAQ 模型是一种用于空气质量预测的模型,该模型包括两部分:WRF(Weather Research and Forecasting model)和 CMAQ(Community Multiscale Air Quality)空气质量模型。WRF 模式是美国国家大气研究中心、美国环境预测中心等科研机构联合开发的一种天气预报模式,相比于传统的天气预报系统,该模式下增添了如气象化学反应机制、垂直扩散效应等物理化学反应机理,通过已知的物化反应机理来对污染物数据进行建模。CMAQ 是美国 1997 年发布的第三代空气质量模型,该模型充分考虑了包括首要污染物、臭氧、空气颗粒物、酸性气体等污染物和源自不同排放源,如电厂、化工厂、交通站等伴随的污染问题,由中尺度气象模式、污染物排放模式和多尺度空气质量模式系统组成,将复杂的空气污染情况进行了综合处理,是目前处于国际领先水平的空气质量模型系统。

但是,现有的 WRF-CMAQ 模型也面临着很多问题。首先,20 世纪末设计的 CMAQ 模型随着新世纪城市化水平和生产力的进一步提升越来越不适用,其研究的模式存在过时

与片面性的问题。而 WRF 模型因为对污染物(如 O₃)的反应机理的研究不够透彻,很多物化反应、中间产物的生成机理还有待进一步研究,因而该系统虽然相比于其他模式有了里程碑式的跨越,但仍有相当的不足,本文旨在通过深度神经网络,绕开目前尚未有重大突破的污染物物化反应机理,基于 WRF-CMAQ 模型进行二次建模,以期提升模型预测准确性(二次建模的想法来自于 2021“华为杯”研究生数学建模竞赛)。

2.2 深度神经网络

随着人工智能领域研究的不断深入,深度神经网络的应用范围也越来越广。深度神经网络^[10-11](Deep Neural Networks, DNN)可用于学习较为复杂的非线性模型,其本质是在感知机上进行的拓展。深度神经网络加入了多层隐藏层,增强了模型的表达能力;输出层的神经元也可以是多个,使得模型能够更加灵活地应对不同的任务,如分类和回归任务,应用范围更加广泛。同时也对激活函数进行了拓展,如本文后续使用的 Sigmoid 函数:

$$f_{\text{SGD}}(z) = \frac{1}{1 + e^{-z}}$$

激活函数的多样性,使得神经网络的表达能力得到进一步增强。

随着神经网络面临的任务不断复杂化,网络结构也愈加复杂,如针对图像分类任务提出的 Resnet 和 Unet 等。本文针对的气象数据仅由数字构成,在数据结构上较为简单,因而选用了较为基础的全连接神经网络。

2.3 联邦学习

为了解决在分布式场景下进行多方训练的隐私保护问题,谷歌提出了联邦学习^[12-15](FL, Federated Learning)架构。联邦学习可以在不传输任何真实数据的情况下,将分布式系统的终端参与到训练中并充分利用其在本地存储的数据。联邦学习将原本集中在中心端的运算量分散到各个终端,传输模型权重而不是原始数据,这样既减少了传输开销和延迟^[16],又保证了用户隐私数据的安全^[14],还分散了在中心端的计算量^[13],一举多得。

在客户端产生的数据是本地存储、本地处理的前提下,联邦学习中客户端和中心服务器定期交换参数。假定输入为矩阵 \mathbf{x}_j^i ,输出为 f_j^i ,在客户端 c_j 上的 loss 函数可以表示为:

$$F_j(\omega, \mathbf{x}_j^1, f_j^1, \dots, \mathbf{x}_j^{D_j}, f_j^{D_j}) = \frac{1}{|D_j|} \sum_{i=1}^{|D_j|} g(\omega, \mathbf{x}_j^i, f_j^i)$$

其中, ω 是客户端 c_j 上和模型相关的参数, D_j 是 c_j 上的数据集。需要着重说明的是, $g(\omega, \mathbf{x}_j^i, f_j^i)$ 是 c_j 上的一个数据块 \mathbf{x}_j^i 的损失计算,而 $F_j(\omega, \mathbf{x}_j^1, f_j^1, \dots, \mathbf{x}_j^{D_j}, f_j^{D_j})$ 是 c_j 上整个本地数据集的损失函数结果。

联邦学习可以获得一个客户端都参与的模型而不用分享任何原始数据。其本质是一个最小化问题:

$$\begin{aligned} \min F(\omega) &\triangleq \sum_{j=1}^n \frac{|D_j|}{|D|} F_j(\omega) \\ &= \frac{1}{|D|} \sum_{j=1}^n \sum_{i=1}^{|D_j|} g(\omega, \mathbf{x}_j^i, f_j^i) \end{aligned}$$

其中, D 代表所有客户端上的数据集的并,即:

$$D = \bigcup_{j=1}^n D_j$$

3 基于联邦学习的污染物浓度二次建模

3.1 数据归一化

由于真实采集的气象数据和 WRF-CMAQ 模型的一次

输出的量纲不同,且不同污染物浓度对空气质量的影响程度也无法仅通过浓度直接比较,因此需要对真实采集的气象数据和 WRF-CMAQ 模型的一次预测数据进行归一化处理。

3.1.1 空气质量指数

在《环境空气质量指数(AQI)技术规定(试行)》中,空气质量指数(AQI)可以判别空气质量等级。计算 AQI 首先需要计算各项污染物的空气质量分指数(IAQI),其计算式如下:

$$IAQI_P = \frac{IAQI_{HI} - IAQI_{Lo}}{BP_{Hi} - BP_{Lo}} \times (C_P - BP_{Lo}) + IAQI_{Lo}$$

其中,IAQI_P代表污染物 P 的空气质量分指数,C_P代表污染物 P 的质量浓度值,BP_{Hi}和BP_{Lo}分别代表与C_P相近的污染物浓度限值的高位值与低位值,而IAQI_{HI}和IAQI_{Lo}则分别代表BP_{Hi}和BP_{Lo}对应的空气质量分指数。各污染物项目浓度限值与空气质量分指数对应表见[环境空气质量指数(AQI)技术规定(试行)]。

将所有采集/预测出的质量浓度转化为空气质量分指数,就可以将所有污染度质量浓度数据转化为无量纲的 IAQI 进行进一步计算与比较,空气质量分指数提供了一种将污染物质量浓度归一化的较为科学的方法(浓度限值与空气质量分指数的对应表根据实际监测情况而来)。

3.1.2 其他数据归一化

除了 6 种本文研究的污染物质量浓度以外,WRF-CMAQ 模型的一次预测与实测的其他气象观测值也需要进行归一化处理。考虑到这些监测值/预测值情况各异,因而采用了简单的最值法进行归一化。即:

$$X_P = X_P / X_{P_{MAX}} \times 500$$

其中,X_P代表气象相关指标 P 的值,X_{P_{MAX}}代表气象相关指标 P 所在检测点的最大值。500 来自于与空气质量分指数 IAQI_P统一的区间跨度。

3.2 深度神经网络

本文研究的污染物浓度预测问题可以看作是一个非线性回归问题,因而采用了 6 层的全连接神经网络,网络结构如表 1 所列。

表 1 网络结构

Table 1 Network structure		
层	输入大小	输出大小
输入层	32	—
隐藏层 1	32	128
隐藏层 2	128	256
隐藏层 3	256	512
隐藏层 4	512	256
隐藏层 5	256	128
隐藏层 6	128	6
输出层	—	6

损失函数采用经典的均方误差(MSE)损失函数:

$$f_{MSE} = \frac{\sum_{i=1}^r (n_i - 1) s_i^2}{N - r}$$

其中, $\sum_{i=1}^r (n_i - 1) s_i^2$ 代表误差平方和,分母 N-r 代表自由度。同时损失函数值也是后文中衡量预测性能/准确度的唯一指标。优化算法采用随机梯度下降(Stochastic Gradient Descent,SGD)算法。

3.3 联邦学习

为了提升多监测站组成的监测系统内的污染物预报准确率,本文引入了联邦学习^[16-18]。联邦学习需要在中心控制

服务器(Server 端)和外部参与节点(Client 端)上进行,本文将各个监测点作为 Client 端,虚构了一个中央预报站作为 Ser-ver 端。数据传输图如图 1 所示。

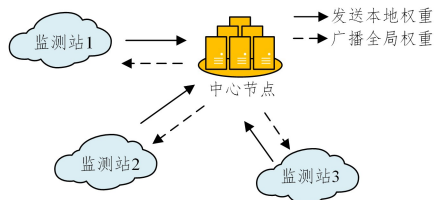


图 1 基于联邦学习的气象观测系统模型(电子版为彩图)

Fig. 1 FL based meteorological observation system

图 1 中,红色虚线标出的是中心节点向各个监测站广播全局权重参数,橙色实线代表各个监测站向中心节点发送自己的本地权重参数。联邦学习中中心节点更新权重的 FedAvg 函数^[19]如下:

$$FedAvg: \omega_{i+1} = \frac{1}{|C|} \sum_{c_j \in C} \omega_{i+1}^{c_j}$$

其中, ω_{i+1} 是第 i 轮迭代的权重参数, $\omega_{i+1}^{c_j}$ 是第 i+1 轮预测站 c_j上的权重参数^[20]。详细的基于联邦学习的气象观测系统算法如算法 1 所示。

算法 1 基于联邦学习的气象观测系统

输入:预测站集合 C,预测站数目 N,初始模型权重 ω_0 ,学习率 η ,每个预测站 c_j上的实测数据集 D_j,迭代论述 E

输出:完成训练的最终预测模型

1. 模型初始化、数据集加载;
2. For(1) i←1 to Edo
3. 中心节点将本轮模型参数 ω_i 广播;
4. For(2) j←1 to N do
5. 预测站 c_j 加载权重参数 ω_i ;
6. 预测站 c_j 根据本地实测数据 D_j,在给定学习率 η 等设定下,进行神经网络训练,获得新的权重参数 ω_j ;
7. 预测站 c_j 将新的权重发送给中心节点;
8. End For(2);
9. 中心节点使用 FedAvg 函数更新权重;
10. End For(1);
11. 输出最终预测模型。

算法包含了训练迭代、预测站双循环,外层循环(1-3、9-10行)是全局模型参数广播^[21]、更新过程,内层循环(4-8行)是各个监测站利用本地实测数据和储存的一次 WRF-CMAQ 一次预测数据进行的本地训练过程。

后续实验会在高相关性(监测点之间距离较近)和无相关性(监测点之间距离极远)两组中进行,同时也会比较这两者之间的性能差异。

4 实验与评估

4.1 数据预处理

每个污染物监测站的数据都包含两部分,WRF-CMAQ 模型的一次预测数据和实测数据(数据来源为 2021 年“华为杯”研究生数学建模竞赛)。其中实测数据包括 6 种污染物的质量浓度、温度、湿度、气压、近地风速、风向,WRF-CMAQ 模型的一次预测数据包括近地 2 米温度(°C)、地表温度(K)、比湿(kg/kg)、湿度(%)、近地 10 米风速(m/s)、近地 10 米风向(°)、雨量(mm)云量、边界层高度(m)、大气压(Kpa)、感热

通量(W/m^2)、潜热通量(W/m^2)、长波辐射(W/m^2)、短波辐射(W/m^2)、地面太阳能辐射(W/m^2)和6种污染物预测的小时平均浓度。

数据中存在部分缺失的情况,对于缺失值我们采用最近一组未缺失该值的数据进行填补。填补完成后对数据进行匹配。我们将时间点 t (精确到小时,下同)的全部实测数据以及时间点 $t+1$ 的WRF-CMAQ一次预测数据作为训练数据,将时间点 $t+1$ 的实测数据作为标签数据,按照时间节点进行匹配,过程中排除所有匹配失败的数据(包含 t 或 $t+1$ 时间的实测数据丢失、时刻 $t+1$ 未产生一次预测数据等情况),将整理后的数据作为最终的数据集。

本文进行测试的2组数据均包含了3个气象监测站,对于每个气象监测站,均选取1000条数据作为测试数据,剩下的数据(每个气象监测点数据量不一)作为训练数据。同时,联邦学习的虚拟中心节点将拥有麾下所有监测站的全部测试数据。

下节将首先针对单个监测点进行WRF-CMAQ模型的一次预报数据精度与二次建模结果进行对比。接下来针对独立监测站组和相关监测站组两组,进行一次预报结果、单个监测站监测结果和联邦学习监测结果的对比。最后总结实验情况。

4.2 单个监测站实验

首先,本文选取了3个独立的(可以理解为监测站之间距离极远,绝对无干扰)监测站,进行二次建模效果的对比实验。3个监测站上的实验结果图如图2—图4所示。

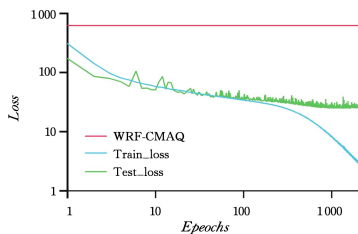


图2 监测点A Loss对比(电子版为彩图)

Fig. 2 Loss comparison of monitoring point A

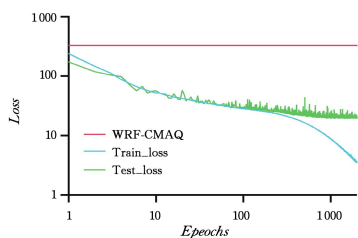


图3 监测点B Loss对比(电子版为彩图)

Fig. 3 Loss comparison of monitoring point B

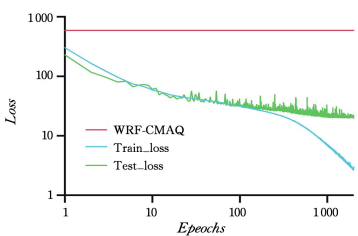


图4 监测点C Loss对比(电子版为彩图)

Fig. 4 Loss comparison of monitoring point C

我们用均方误差损失函数值作为性能衡量依据。图2—图4中,红线为每个监测点全数据集的WRF-CMAQ一次

预测的Loss值,蓝线为训练数据集的Loss值,绿线为测试数据集的Loss值。训练1000个循环后,训练平均MSE Loss普遍在5以下,监测点A,B,C上的测试数据集的Loss分别缩小到了24.56,19.10,19.66,相比WRF-CMAQ的一次Loss值(624.22,328.76,591.09)分别缩小到了3.93%,7.97%,6.47%。

4.3 独立监测站组实验

接下来,我们在地理上独立的监测站A,B,C以及虚拟的中心节点上进行训练与测试。中心节点和各个独立监测节点采用同样的测试集(每个独立监测站各提供1000条不重复的随机测试数据)。我们将中心节点和各个监测站进行联邦学习的结果与各个监测站单独训练数据的性能进行了对比,如图5所示。WRF-CMAQ的一次预测结果loss仍然远高于所有采用深度学习神经网络的其他对比组。使用联邦学习算法的训练效果比监测站A,B,C分别提升了59.78%,68.89%和53.88%(计算方法为loss差值/对比组原loss值)。

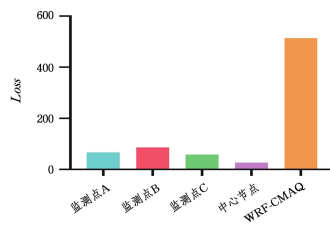


图5 独立监测点组 Loss对比

Fig. 5 Loss comparison of independent monitoring group

4.4 相关监测站组实验

最后,我们在一组高相关性的数据上进行实验。监测站 A_1, A_2, A_3 是一组在同一区域且地理距离无法排除其相互干扰性的监测站组。同样的,我们也虚构了一个该区域的中心监测节点且中心节点和各个独立监测节点采用同样的测试集(每个独立监测站各提供1000条不重复的随机测试数据)。我们将训练后的模型性能的测试结果采用与4.2节中相同的方式进行了对比。如图6所示。经过计算,使用联邦学习的算法的训练效果相比监测站 A_1, A_2, A_3 分别提升了63.20%,46.79%和61.75%。

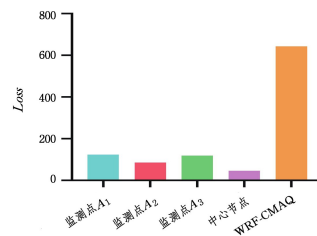


图6 相关监测点组 Loss对比

Fig. 6 Loss comparison of related monitoring group

4.5 实验结果总结

本节的实验证明了采用深度神经网络对WRF-CMAQ模型的一次预测结果进行二次建模是十分有效的,所有的模型测试都显示,采用二次建模的效果远远优于一次建模的结果。同时,在有限考虑模型泛化能力的情况下,采用联邦学习的方案在模型泛化上高于单一监测站(实验结果表明在混合测试集上联邦学习方案取得了最优成果,但在单个监测站点可能并没有获得最优解)。实验中,相比于一次WRF-CMAQ的一次预测结果,深度神经网络的方案在均方误差上最多

缩小到了 3.93%。同时,采用联邦学习的方案相比于单个监测站点最多提升了 68.89%的性能。

结束语 本文采用了深度神经网络对 WRF-CMAQ 模型的一次污染物浓度预测输出进行二次建模,来提升模型输出与真实值之间的误差,进而提升污染物预报的准确率。实验结果表明,深度神经网络可以将预测值与实测值之间的 MSE 损失值降低到 3.93%,同时联邦学习可以收获泛化性更强的预测模型,基本满足了在仅保有少量监测点的数据区域内进行预测的需求。本文讨论的模型预测时间相比于 WRF-CMAQ 模型预测时间较短,仅讨论了第一次预测的情况,这对于预报污染物浓度高的场景是有帮助的,但对长时间段预测的研究还不足(因为 WRF-CMAQ 模型中对气象情况的模拟可以提升长时间段的预测效果),这也是后续工作开展的方向。

未来可以从几个方面展开进一步研究:1)使用更加复杂的网络和精巧的算法,进一步提升模型预测能力;2)在采用联邦学习算法的同时,设计更科学的权重分配方案,在保证模型泛化能力的同时,进一步提升模型在单个需求预测点的性能;3)对 WRF-CMAQ 模型本身进行优化,加入最新的物理化学大气等基础学科研究成果,提升二次建模的输入数据精确度。

参 考 文 献

[1] HU J, CHEN J, YING Q, et al. 1 One-Year Simulation of Ozone and Particulate Matter in China 2 Using WRF/CMAQ Modeling System[J]. *Atmospheric Chemistry & Physics Discussions*, 2016, 16(16):1-31.

[2] HU J, LI X, HUANG L, et al. Ensemble Prediction of Air Quality Using the WRF/CMAQ Model System for Health Effect Studies in China[J]. *Atmospheric Chemistry and Physics*, 2017, 17(21):13103-13118.

[3] STEYN D, CHAUMERLIAC N. Recent Advances in Modeling of the Atmospheric Boundary Layer and Land Surface in the Coupled WRF-CMAQ Model[M]. Berlin: Springer International Publishing, 2016.

[4] PLEIM J, GILLIAM R, APPEL W, et al. Application and Evaluation of High-Resolution WRF-CMAQ with Simple Urban Parameterization [M]. Berlin: Springer International Publishing, 2014.

[5] BNW A, BHG A, CF J, et al. Simulation of Ozone Formation at Different Elevations in Mountainous Area of Hong Kong Using WRF-CMAQ Model[J]. *Science of The Total Environment*, 2015, 505(118):939-951.

[6] LEE Y, GERRITY D, LEE M, et al. Prediction of Micropollutant Elimination during Ozonation of Municipal Wastewater Effluents: Use of Kinetic and Water Specific Information[J]. *Environmental Science & Technology*, 2013, 47(11):5872-5881.

[7] MANIAS D M, SHAMI A. Making a Case for Federated Learning in the Internet of Vehicles and Intelligent Transportation Systems[J]. *IEEE Network*, 2021, 35(3):88-94.

[8] 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.

[9] BROWN D S, FLAGG E. Empirical Prediction of Organic Pollutant Sorption in Natural Sediments[J]. *Journal of Environmental Quality*, 1981, 10(3):382-386.

[10] ZHANG W, LI X, MA H, et al. Federated Learning for Machine-

ry Fault Diagnosis with Dynamic Validation and Self-supervision [J]. *Knowledge-Based Systems*, 2021, 213(1):106679.

[11] MING C, YIN Y, MIN X. Prediction of Hourly Air Pollutant Concentrations near Urban Arterials Using Artificial Neural Network Approach[J]. *Transportation Research Part D*, 2009, 14(1):32-41.

[12] HINTON G, DENG L, YU D, et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups[J]. *IEEE Signal Processing Magazine*, 2012, 29(6):82-97.

[13] CIRESAN D, MEIER U, SCHMIDHUBER J. Multi-column Deep Neural Networks for Image Classification[C]// *Computer Vision & Pattern Recognition. IEEE*, 2012.

[14] EPHRAI M. Speech Enhancement Using a Minimum Mean Square Error Short-time Spectral Amplitude Estimator [J]. *IEEE Trans Acoust Speech Signal Process*, 1984, 32(6):1109-1121.

[15] KONEN J, MCMAHAM H, YU F, et al. Federated Learning: Strategies for Improving Communication Efficiency[J]. *arXiv*: 1610.05492, 2016.

[16] SAMARAKOON S, BENNIS M, SAAD W, et al. Distributed Federated Learning for Ultra-Reliable Low-Latency Vehicular Communications[J]. *IEEE Transactions on Communications*, 2020, 68(2):1146-1159.

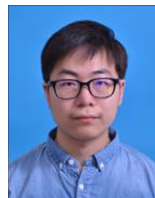
[17] REHAK D R, DODDS P, LANNOM L. A Model and Infrastructure for Federated Learning Content Repositories[C]// *Proceedings of the 1st International Workshop on Interoperability of Web-based Educational Systems*, 2005:3-12.

[18] THEODORAS B, CHENRD, MELA T, et al. Federated Learning of Predictive Models from Federated Electronic Health Records[J]. *International Journal of Medical Informatics*, 2018, 112:59-67.

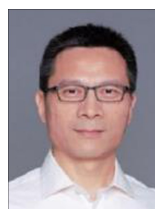
[19] SUBRAMANYA T, RIGGIO R. Centralized and Federated Learning for Predictive VNF Autoscaling in Multi-domain 5G Networks and Beyond[J]. *IEEE Transactions on Network and Service Management*, 2021, 18(1):63-78.

[20] HUR S K, HO C H, KIM J, et al. Systematic Bias of WRF-CMAQ PM10 Simulations for Seoul, Korea[J]. *Atmospheric Environment*, 2021, 244:117904.

[21] CHEN D, XIA L, GUO X, et al. Impact of inter-annual meteorological variation from 2001 to 2015 on the contribution of regional transport to PM_{2.5} in Beijing, China[J]. *Atmospheric Environment*, 2021(D22):118545.



QIAN Dong-wei, born in 1998, postgraduate, is a member of China Computer Federation. His main research interests include federated learning and semi-supervised learning.



WEI Tong-quan, born in 1973, Ph.D, associate professor. His main research interests include edge computing, cloud computing, and design automation of intelligent systems and cyber physical systems.