2021 IEEE International Conference on Big Data (Big Data)

Spatially-distributed Federated Learning of Convolutional Recurrent Neural Networks for Air Pollution Prediction

Do-Van Nguyen, Koji Zettsu
Big Data Integration Research Center
National Institute of Information and Communications Technology (NICT)
Tokyo, Japan
ngdovan@nict.go.jp, zettsu@nict.go.jp

2021 IEEE International Conference on Big Data (Big Data) | 978-1-6654-3902-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/BigData52589.2021.9671336

***Abstract*—Air pollution prediction for smart city applications has been attracted in artificial intelligence research to overcome problems to the health of citizen. Conventionally, environmental IoT data is gathered from monitoring station sensors then is sent to servers for centralized predictive model training at a whole region. That causes latency issues in data transmission from IoT devices to cloud servers. This paper describes federated learning paradigm approach for air pollution prediction model training on environmental monitoring sensor data. In the research, we design distributed learning framework that assists cooperative training among participants from different spatial areas such as cities and prefectures. At each area, Convolutional Recurrent Neural Networks (CRNN) are trained locally aiming to predict local Oxidant warning level while aggregated global model enhances distilled knowledge from all areas of a region. The research illustrates that designed common parts of CRNN can be fused globally meanwhile adaptive structure at predictive part of the deep neural network model can capture different environmental monitoring stations configuration at local areas. Some experiment results also hint methods to keep balance between federated learning synchronous training rounds and local deep neural network training epochs to maximize accuracy of the whole federated learning system. The results also prove that new participating areas can train and quickly obtain optimized local models by using transferred common global model.**

***Index Terms*—Internet of Things (IoT), Environmental Sensors, Spatial-Temporal Data, Federated Learning, Air Pollution Prediction.**

I. INTRODUCTION

It is undoubted that outdoor particulate air pollution $(PM_{2.5})$ is one of the main factors that cause mortality risk, thus is associated with increasing in global disease treatment expense. Besides, other pollution factors including PM_{10}, SO_2, NO_2, O_3 are also concerned by governments and organizations. It is noticed that Oxidant gases have strong impact on relationship between $PM_{2.5}$ and non-accidental, cardiovascular and respiratory mortality [13].

Recently, numerous studies have been conducted in machine learning research to build predictive models from environmental IoT sensor data, that can be utilized for forecasting air pollution in smart city applications [7]. Studies in Zhao et. al. [16], [17] suggest that environmental and meteorological data can be combined and transformed to the matrix representation, then fed into conventional convolutional recurrent neural network [16], [17] or directly used as graph structure [18] to predict multiple outcomes of air pollution. With the same approach of spatial transformation, Yi et. al. [15] employed embedding model to encode multi-modalities before combining into a fusion weighted network to predict Air Quality Index (AQI). Research in [2] analyzes important factors that cause air pollution and put into a special customized design of encoder-decoder framework with CNN-LSTM to generate heat-map of future air quality. Study in [11] is also to name other empirical experiments with CNN-LSTM to encode spatial and temporal in air pollution prediction.

Traditionally, machine learning models usually are trained with centralized data set which heavily require data transmission between IoT data collected sides and AI servers. The concept of Federated Learning (FL) was first proposed by McMahan et. al. [1], [8], [12] which enables cooperative training among participants those have their own data corpus. Training without exchange private data thus reduces transmission bandwidth as well as keeps personal data secured. To leverage the advantage of federated learning, research in [4] shows a design of distributed inference framework for urban environment sensing. Study in [10] demonstrates using federated learning to help unmanned aerial vehicles achieving accurate AQI prediction.

While most of conventional federated learning approach in air pollution prediction study on optimizing transferring protocol between devices/edges and servers, this research will focus on heterogeneous data issues, in which data space maybe different among edge sides of geographical areas. Leveraging CRNN studies on air pollution prediction [16], [17], we contribute a study to design and develop federated learning framework on air pollution prediction with an illustration on oxidant warning problem.

Our main contribution can be summarized as follows:

- We design a federated learning system that can capture sensor data at monitoring stations of spatial areas including cities and prefectures, train predictive models at edges of cities/prefectures and share models to others to transfer

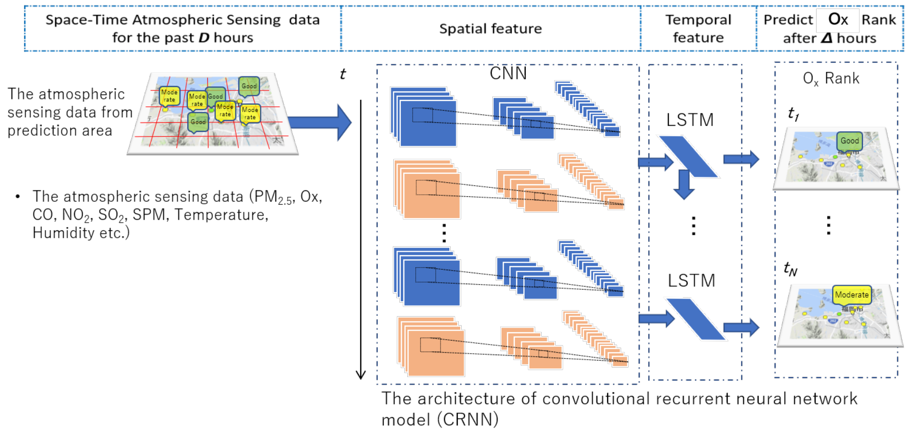

Figure 1. Air quality prediction using Convolutional Recurrent Neural Network.

knowledge meanwhile eliminate IoT data transmission in between the whole region.

- We propose Federated Learning CRNN Model for air pollution prediction. The model contains localized neural network layers, which have different structures among edges, and common neural network layers to enable sharing knowledge on federated learning framework.
- We build spatial averaging aggregation function of federated learning paradigm and prove the concept through experimental design on actual environmental and meteorological data set of Kanto region, Japan.
- We empirically show that choosing appropriate configuration of federated training rounds and local training epochs can lead to a higher accuracy of general models. We also illustrate knowledge gained from federated learning on a set of participants can be transferred to faster local model training at a new participant.

It is noticed that communication [3], [19], model attack [20] and other issues of federated learning are not in the scope of this study.

The remainder of this paper is organized as follows. Section II will discuss some related works to discover the conventional research in air quality prediction and federated learning. Next, Section III will describe novel design of local CRNN model and federated learning framework for oxidant prediction. Consequently, experiment setup, results and discussion will be in Section IV. Lastly, Section V will summarize the research and open future works.

II. Research Background

This section will describe some related works on using deep convolutional recurrent network to predict AQI pollution and some federated learning techniques for intelligent IoT application to prepare for the later proposed methodology.

A. CRNN Models for Air Quality Prediction in Smart Cities

The local models are CRNN structure to predict air pollution after a period. First, spatial-temporal atmospheric sensing data from all environmental monitoring stations in a city/prefecture is transformed into spatial feature in the form of matrix representation, which can be fed into deep learning model using convolutional neural network (CNN) [9] to classify air pollution levels. To model the link between time scale in the period, a Long Short-Term Memory (LSTM) [6] is used for sequential structure encode which capture recurrent information over time. Thus, using t-cell of LSTM, the model can predict air pollution in a t-time steps. The Figure 1 illustrates overall diagram of a local CRNN model for ranking Oxidant after several hours.

CRNN model has been successfully applied to predict air pollution in [15], [17], [18] where input of the CRNN are space-series of the r sensor s in sensor data space $S = S_1, ..., S_r$. The sensor values coordinated with longitude and latitude are interpolated and transformed into matrix representation $I_{m \times n \times r}$ by inverse distance weighting (IDW) [5] before feeding into CNN. Number of sensor types in environmental and meteorological data forms r channels of $m \times n$ input to convolutional layers of CNN. The output of CNN are connected to input of LSTM to predict air pollution over periods such that $rank(O_x) = softmax(LSTM(CNN(I)))$.

B. Federated Learning for Intelligent IoT Applications

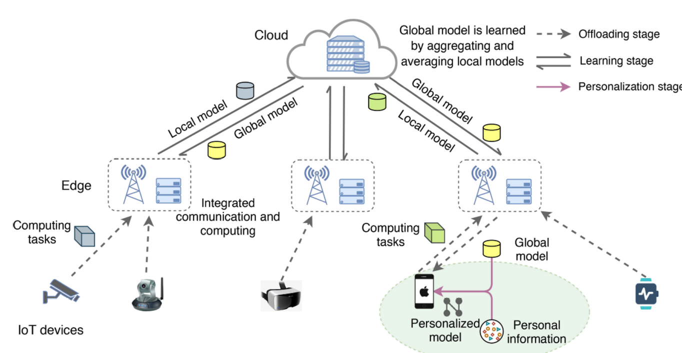

Figure 2. The federated framework for intelligent IoT applications [14].

Federated learning is proposed to train global machine learning model by aggregating local model trained locally on massive IoT data samples to keep personal data privacy and reduce data transmission. Figure 2 illustrates a complex IoT environment [14] in cloud-edge architecture for intelligent IoT applications in which distributed data from smart devices and IoT with limited computational are sent to Edge (of smart homes, smart cities for example) for local model training. At central cloud server, global model will be aggregated from edges with federated learning paradigm without data exchange among the edges.

In a federated learning system, there are a set of K participants $P = P_1, ..., P_K$ that own their private datasets $D_{k\in[1,K]}$. In an iterative training round, a global model w_G is optimized from many local models w_k trained locally on local private datasets. Therefore, the objective of learning problem is to find an optimal aggregate function f such that:

$$w_G = f(w_k). \tag{1}$$

The optimizing process is thus to minimize

$$\min_{w_G} f(w_k|w_G) = \sum_{k\in[1,K]} p_k F_k(w_k) = \mathbb{E}_k[F_k(w_k)], \tag{2}$$

where $p_k \geq 0$ is the data D_k probability distribution. In general, $F_k(w_k) := \mathbb{E}_{x_k\sim D_k}[f_k(w_k|x_k)]$ and $p_k = n_k/n$, empirically, where n_k is the cardinality of data sample available at participant k and $n = \sum_k n_k$ is the total number of the whole data corpus over the all participants at the same training round.

Conventional study in federated learning shows that Federated Averaging Aggregation (FedAvg) [12] works effectively with heterogeneous data corpus. In FedAvg, local optimizer is stochastic gradient descent (SGD) with the same hyper-parameters including learning rate and local epochs on each participant to optimize Eq. (2). At a training round t, a selected number of participants train locally then send local models to server to update global model using averaging aggregate function f in Eq. (1) by:

$$w_G^t = \sum_k p_k w_k^t = \frac{1}{n}\sum n_k w_k^t. \tag{3}$$

The process of federated learning over T times of training rounds with federated averaging aggregation can be summarized as Alg. 1.

III. Methodology

In this section, we first show the local CRNN structure for oxidant warning prediction. This model structure is designed toward to be cooperatively trained among participating areas. Meanwhile, it can be adapted with heterogeneous monitoring environmental local data. Then we propose federated learning framework with spatial averaging aggregation approach to fuse the common parts of CRNN models at server side.

Algorithm 1 Federated learning

```
Initialize global model w_G^(0)
for each round t = 0, 1, 2 ... T do
    Server select a subset of participants
    Server distributes the global model w_G^t the selected participants: w_k^t = w_G^t.
    Each selected participant k updates local model w_k^t for E epochs to obtain w_k^{t+1}
    Each participant send w_k^{t+1} back to the server.
    The server update global model w_G^{t+1} using Eq. 3.
end for
```

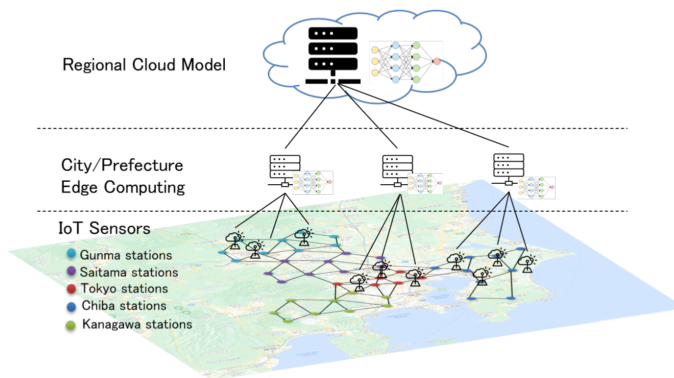

Figure 3. Spatially-distributed Federated Learning of Regional AQI Pollution Prediction System.

A. Local CRNN model

The Federated learning CRNN model to predict oxidant warning level in this research is similar to CRNN model in [17]. However, different from the predecessor, instead of using trans-boundary data from other areas, local CRNN models are trained on only local monitoring data, but such models may benefit from knowledge shared from other location. As output, CRNN models will predict oxidant level at all monitoring stations in the trained local area. Therefore, a local model of CRNN w_k is defined by $w_k := \{w_{k,c}, w_{k,l}\}$, where $w_{k,c}$ and $w_{k,l}$ is (common) generalized model and localized model of a CRNN, respectively.

In this research, we consider a case such that the global server also has its own public data. Hence, at server side, it likewise has a CRNN model $w_G := \{w_{G,c}, w_{G,l}\}$ trained on central public data. We will discuss global utilized CRNN models in the experiment section.

In more detail, a CNN part of CRNN consists of common convolutional layers which can be shared among different areas (note that participants may have different number of stations) and localized dense layers which represent station distribution in each local area. Each output of the dense layers predicts air pollution outcome of a station. We also use localized layers as auxiliary models in order to pre-train CNN, thus it can capture spatial information before feeding into LSTM. With the same structure objective design, LSTM part also has generalized LSTM cells and localized dense layers

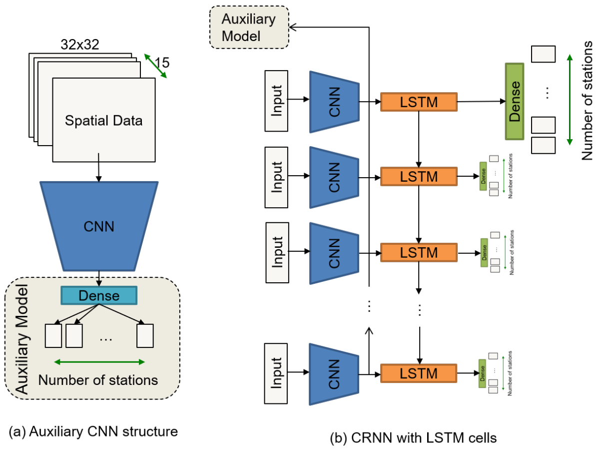

Figure 4. CRNN-Structure.

for common and localization, respectively. The illustration of neural network structure can be found in Fig. 4.

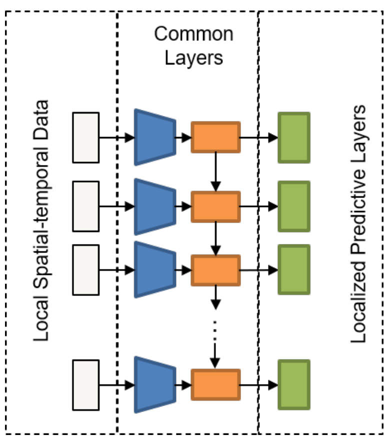

Figure 5. Layer Split of CRNN in Federated Learning

In federated learning setup, the whole CRNN of a local area can be trained privately using local data. However, only common layers (Fig. 5) will be feedback/transferred between clients and servers for aggregation. The next section will give discussions on aggregation methods which combine those common layers at server side.

B. Federated Learning Aggregation Methods

This section will explain how we design aggregation functions in Eq. (1) of global model $w_{G,c}$.

Fig. 6 illustrates aggregation methods for fusion model trained on spatial data. In the averaging approach, the global common model contains all information from the whole region, such that:

$$w_{G,c} = \sum_{\forall k \in K} p_k w_{k,c}, \tag{4}$$

will be transferred to all areas in the region.

The algorithms for CRNN federated learning can be found in Alg. 2. Note that in using CRNN structure, all participants

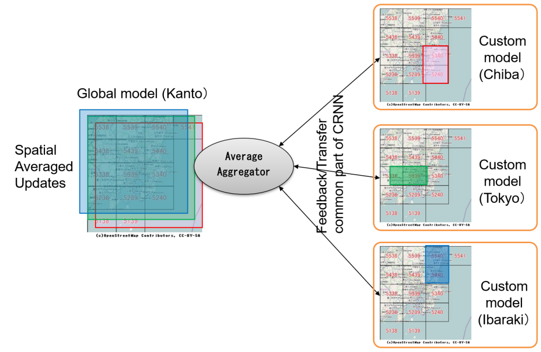

Figure 6. Spatial Averaging Aggregation

Algorithm 2 CRNN Averaging Aggregation FL Algorithm

Initialize global model $w_G^{(0)} = \{w_{G,c}^{(0)}, w_{G,l}^{(0)}\}$
Initialize local CRNN model for all participants
for each round t = 0, 1, 2 . . . T **do**
  Server select a subset of participants
  Server distributes the global common model $w_{G,c}^t$ the selected participants: $w_{k,c}^t = w_{G,c}^t$.
  Each selected participant k updates local model w_k^t for E epochs to obtain $w_k^{t+1} = \{w_{k,c}^{t+1}, w_{k,l}^{t+1}\}$
  Each participant send $w_{k,c}^{t+1}$ back to the server.
  The server update global model w_G^{t+1} using Eq. 4.
end for

also must initialize their local models due to localization parts of CRNN are created based on monitoring station distribution within the area.

IV. Experiment, Results and Analysis

A. Local environment data and experimental result illustration

Data for the experiments is collected from all prefectures in Kanto region, Japan, including Chiba, Gunma, Ibaraki, Kanagawa, Saitama, Tochigi, and Tokyo, from 2018 to 2021, by The Atmospheric Environmental Regional Observation System (AEROS) [1]. Fig. 7 shows the spatial distribution of stations in each prefecture in different color on the map of the region. Detail attribute description of the atmospheric sensing data can be found in Table I.

In the experiment, we show empirical studies of CRNN federated learning framework for Oxidant warning prediction. Different local CRNN models will be used for different ranking of oxidant warning including rank 3, 4 and 5. We first analyze the observed training convergence of federated learning system over number of training rounds and measure accuracy of generalized model. Then we discover the ability of transferring knowledge to untrained local models of newly added participants.

Fig. 8 to Fig.13 depict learning curves of local training at participants and transferring to global/new participant models. In the figures, the train curves represent accuracy calculated

[1]http://soramame.taiki.go.jp/

Table I
THE ATMOSPHERIC SENSING DATA ATTRIBUTES

Atmospheric sensing data (hourly measurement)	Unit
SO_2	ppm
NO_x	ppm
NO	ppm
NO_2	ppm
CO	ppm
O_x	ppm
NMHC	ppm
CH_4	ppm
THC	ppm
STM	mg/m^3
$PM_{2.5}$	μg/m^3
Wind direction	°
Wind speed	m/s
Temperature	°C
Humidity	%

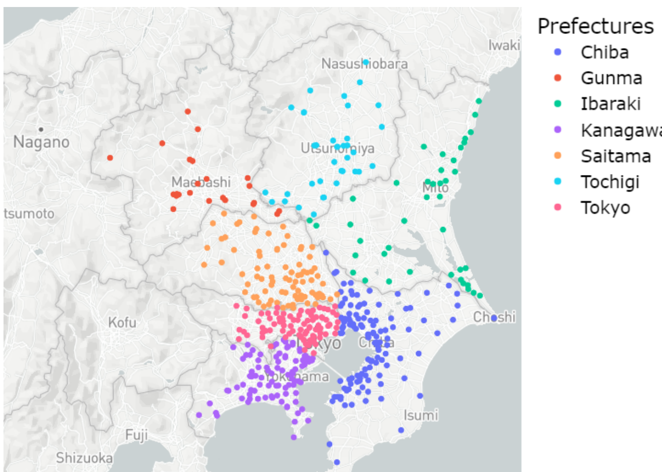

Figure 7. Environmental Monitoring Stations Distribution.

from the local model on training datasets to illustrate how well CRNN models are learnt. The validation curves show accuracy calculated from a hold-out validation datasets to evaluate CRNN models after training or fine-tuning.

B. Local training and global model improvement analysis

In the first experiment, we train CRNN model on datasets of some prefectures to observe the training curves in different setting including:

- *Configuration 1*: Using 50 rounds of federated learning with 2 epochs in each local training.
- *Configuration 2*: Using 20 rounds of federated learning with 5 epochs in each local training.

In *Config.1*, as can be seen from Fig. 8 for a local training (Chiba prefecture to illustrate), due to only low number of epochs in each local training round, the accuracy has small increasing after each round. It then drops sharply after aggregated and transferred from the global model. After total 100 epochs of 50 rounds, the training curve seems to be converged.

At the server-site, global model accuracy increases noticeably for rank 3 oxidant prediction. However, it notices the worst accuracy for rank 4 and rank 5. Therefore, too little training epochs at participants may not capture and transfer knowledge to the server even it shows convergence in local training curves.

In the *Config.2*, it shows the same pattern with the first configuration, training curves fluctuate along each training round. It goes down to the lowest point after aggregation and reaches to higher stages after local adaptation.

On the other hand, on the opposite of the first configuration, the pattern of global model gradually increases after number of rounds and training epochs for the three ranking models. It suggests that local training epochs should be large enough to fully capture knowledge from each local side and be able to transfer to the server for aggregation.

C. Local model improvement by transfer learning in comparison with training from scratch.

In this experiment, we add a new area that is not included in the set of participants in the previous federated training experiment. Suppose that we transfer the common global model trained in the previous experiment to a prefecture (Saitama for example), we do fine-tuning using local datasets and compare learning curves with local training from scratch.

Fig. 12 and Fig. 13 show that both fine-tuning from common global models and training from scratch, respectively, converge after number of epochs. Transferring from aggregated global models seems to be converged as the same speed as local models trained from scratch. It proves a potential use of transfer learning for air pollution prediction within federated learning paradigm, hence it reduces data transmission and protects data privacy.

V. CONCLUSION

This research summaries a federated learning approach to air pollution prediction in smart city applications and presents empirical studies on environmental IoT data using convolutional recurrent neural network models on oxidant warning model. The results shows that CRNN models can capture spatial-temporal local information and be able to share knowledge among participating cities by leveraging federated learning methodology. Therefore, transferring models instead of exchanging data, federated learning is likely to eliminate data collecting from monitoring edges to central servers. The results also illustrate that new participating cities/prefectures can use global shared model to accelerate their local training even those cities/prefectures have not contributed to train global models.

Current averaging aggregate function leverages knowledge generalized from all locations then transfers to all participants without caring spatial relationship among areas. Future work of this research aims to customize feedback models from central server to participants. Taking neighbouring update aggregation for example, local models will be updated with the most related spatial information. Thus, a local model accuracy is expected to increase with shared knowledge from its most related neighbors.

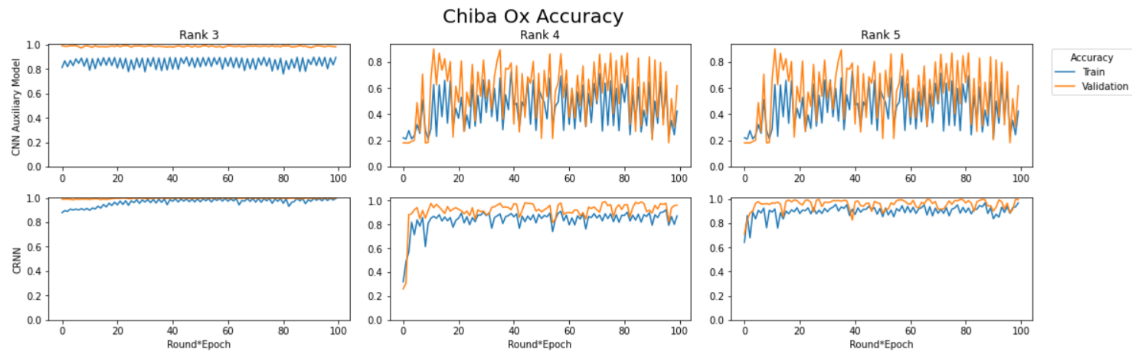

Figure 8. Training curves at locals in the configuration of 50 round - 2 epochs.

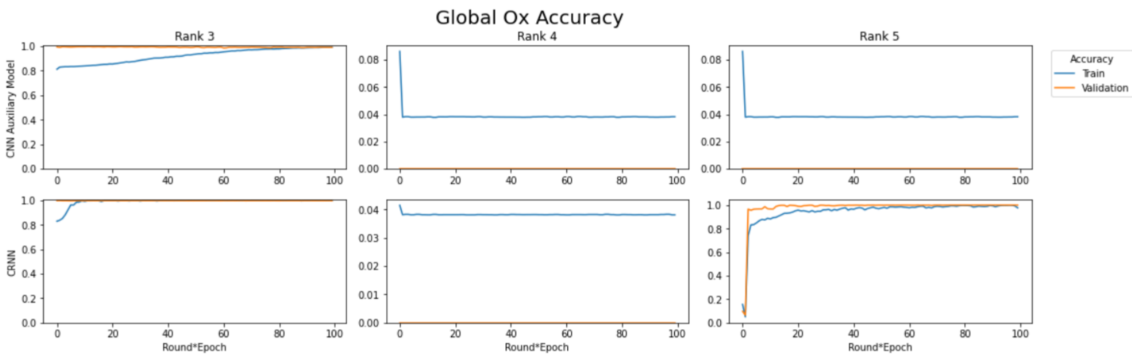

Figure 9. Global accuracy after aggregation in the configuration of 50 round - 2 epochs.

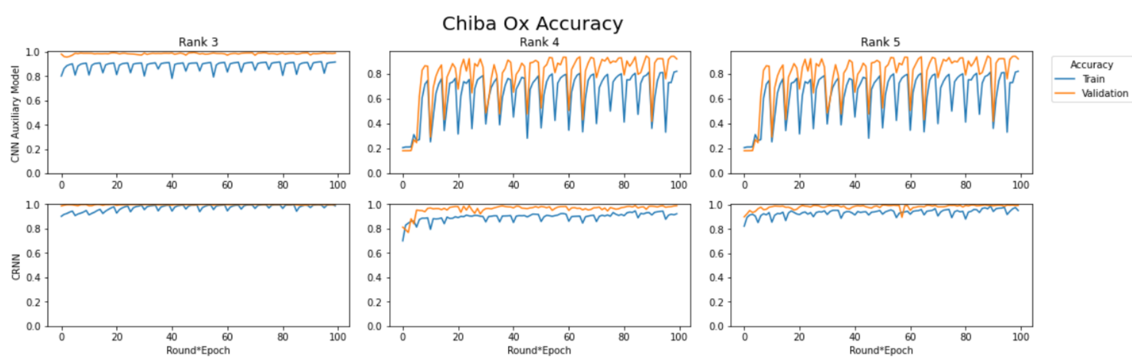

Figure 10. Training curves at locals in the configuration of 20 round - 5 epochs.

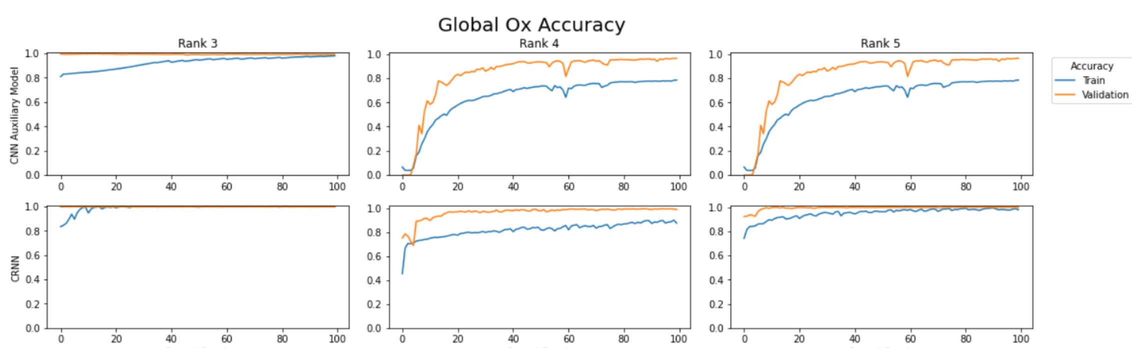

Figure 11. Global accuracy after aggregation in the configuration of 20 round - 5 epochs.

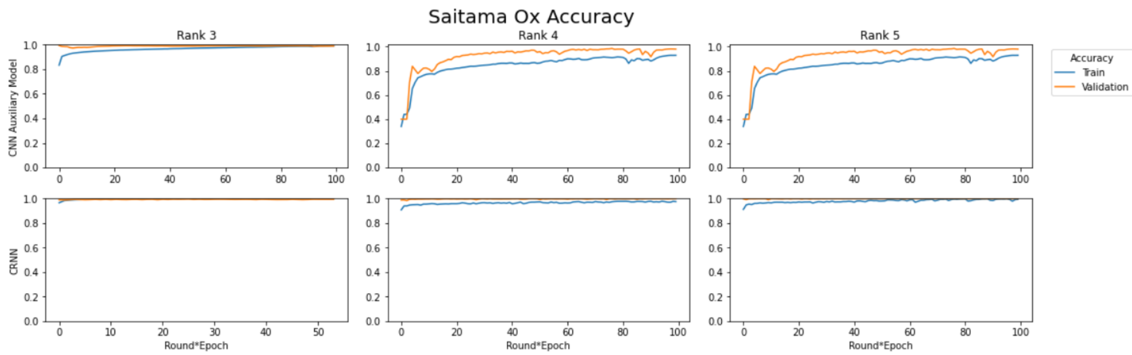

Figure 12. Fine-tuning at local sites using common global shared model.

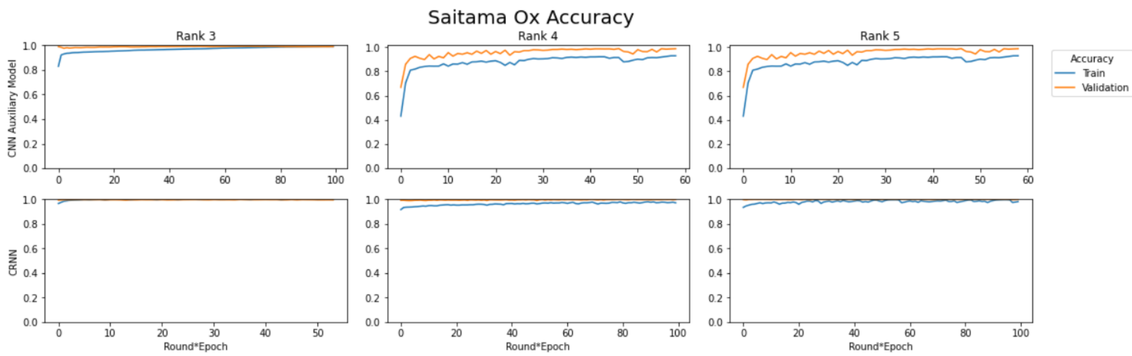

Figure 13. Local training from scratch at a local side.

REFERENCES

[1] Keith Bonawitz, Hubert Eichner, Wolfgang Grieskamp, Dzmitry Huba, Alex Ingerman, Vladimir Ivanov, Chloe Kiddon, Jakub Konečný, Stefano Mazzocchi, H. Brendan McMahan, Timon Van Overveldt, David Petrou, Daniel Ramage, Jason Roselander, "Towards Federated Learning at Scale: System Design", arXiv:1902.01046v2 [cs.LG]

[2] Tien-Cuong Bui, Joonyoung Kim, Taewoo Kang, Donghyeon Lee, Junyoung Choi, Insoon Yang, Kyomin Jung, Sang Kyun Cha, "STAR: Spatio-Temporal Prediction of Air Quality Using A Multimodal Approach", Electrical Engineering and Systems Science

[3] Feng Yin, Zhidi Lin, Yue Xu, Qinglei Kong, Deshi Li, Sergios Theodoridis, Shuguang (Robert) Cui "FedLoc: Federated Learning Framework for Data-Driven Cooperative Localization and Location Data Processing," IEEE Open Journal of Signal Processing, November 6, 2020

[4] Yujia Gao , Liang Liu, Binxuan Hu, Tianzi Lei, and Huadong Ma, "Federated Region-Learning for Environment Sensing in Edge Computing System," in IEEE Transactions on Network Science and Engineering, vol. 7, no. 4, pp. 2192-2204, 1 Oct.-Dec. 2020, doi: 10.1109/TNSE.2020.3016035.

[5] George Y. Lu, David W. Wong, "An adaptive inverse-distance weighting spatial interpolation technique." Computers & Geosciences, 34,9 (2008),1044-1055.

[6] S. Hochreiter and J. Schmidhuber, "Long short-term memory." Neural Computation 9.8 1997:p. 1735–1780.

[7] Ditsuhi Iskandaryan, Francisco Ramos and Sergio Trilles:"Air Quality Prediction in Smart Cities Using Machine Learning Technologies Based on Sensor Data: A Review," Appl. Sci. 2020, 10(7), 2401.

[8] Peter Kairouz; H. Brendan McMahan; Brendan Avent; Aurélien Bellet; Mehdi Bennis; Arjun Nitin Bhagoji; Kallista Bonawit; Zachary Charles; Graham Cormode; Rachel Cummings; Rafael G. L. D'Oliveira; Hubert Eichner; Salim El Rouayheb; David Evans; Josh Gardner; Zachary Garrett; Adrià Gascón; Badih Ghazi; Phillip B. Gibbons; Marco Gruteser; Zaid Harchaoui; Chaoyang He; Lie He; Zhouyuan Huo; Ben Hutchinson; Justin Hsu; Martin Jaggi; Tara Javidi; Gauri Joshi; Mikhail Khodak; Jakub Konecný; Aleksandra Korolova; Farinaz Koushanfar; Sanmi Koyejo; Tancrède Lepoint; Yang Liu; Prateek Mittal; Mehryar Mohri; Richard Nock; Ayfer Özgür; Rasmus Pagh; Hang Qi; Daniel Ramage; Ramesh Raskar; Mariana Raykova; Dawn Song; Weikang Song; Sebastian U. Stich; Ziteng Sun; Ananda Theertha Suresh; Florian Tramèr; Praneeth Vepakomma; Jianyu Wang; Li Xiong; Zheng Xu; Qiang Yang; Felix X. Yu; Han Yu; Sen Zhao, Advances and Open Problems in Federated Learning , Now Foundations and Trends, 2021.

[9] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks." In: Advances in Neural Information Processing Systems. 2012. p. 1097–1105.

[10] Yi Liu, Jiangtian Nie, Xuandi Li, Syed Hassan Ahmed, Wei Yang Bryan Lim, and Chunyan Miao. Federated learning in the sky: Aerial-ground air quality sensing framework with uav swarms. IEEE Internet of Things Journal, 8(12):9827-9837, 2021.

[11] V. Le, T. Bui and S. Cha, "Spatiotemporal Deep Learning Model for Citywide Air Pollution Interpolation and Prediction," 2020 IEEE International Conference on Big Data and Smart Computing (BigComp), 2020, pp. 55-62, doi: 10.1109/BigComp48618.2020.00-99.

[12] McMahan, H. B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B. A. Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, pp. 1273–1282, 2017

[13] Weichenthal, S., Pinault, L.L. & Burnett, R.T. Impact of Oxidant Gases on the Relationship between Outdoor Fine Particulate Air Pollution and Nonaccidental, Cardiovascular, and Respiratory Mortality. Sci Rep 7, 16401 (2017). https://doi.org/10.1038/s41598-017-16770-y

[14] Qiong Wu, Kaiwen He, Xu Chen, "Personalized Federated Learning for Intelligent IoT Applications: A Cloud-Edge Based Framework," IEEE Open Journal of the Computer Society, Vol.1 , 2020.

[15] Xiuwen Yi, Junbo Zhang, Zhaoyuan Wang, Tianrui Li, Yu Zheng, "Deep Distributed Fusion Network for Air Quality Prediction". In Proceed- ings of KDD'18, London, United Kingdom, August 19-23, 2019

[16] Peijiang Zhao and Koji Zettsu: "Convolution Recurrent Neural Networks for Short-Term Prediction of Atmospheric Sensing Data," 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData),

2018, pp. 815-821, doi: 10.1109/Cybermatics2018.2018.00159.
[17] Zhao, Peijiang and Zettsu, Koji. (2019). Convolution Recurrent Neural Networks Based Dynamic Transboundary Air Pollution Predictiona. 410-413. 10.1109/ICBDA.2019.8712835.
[18] P. Zhao and K. Zettsu, "MASTGN: Multi-Attention Spatio-Temporal Graph Networks for Air Pollution Prediction," 2020 IEEE International Conference on Big Data (Big Data), 2020, pp. 1442-1448, doi: 10.1109/BigData50022.2020.9378156.
[19] Zhenyi Lin, Xiaoyang Li, Vincent K. N. Lau, Yi Gong, Kaibin Huang: "Deploying Federated Learning in Large-Scale Cellular Networks: Spatial Convergence Analysis", arXiv:2103.06056v1 [cs.IT]
[20] Zhuohang Li, Luyang Liu, Jiaxin Zhang, Jian Liu: "Byzantine-robust Federated Learning through Spatial-temporal Analysis of Local Model Updates", arXiv:2107.01477v2 [cs.LG]