Implementation and Validation of an Air Quality Prediction Model Based on Federated Learning

Jiaxuan Xu

School of Economics and Management, Harbin Institute of Technology, Harbin, China

qing3302294359@outlook.com

Abstract. Ground ozone and other air pollutants pose a great threat to the health. Because the data of different monitoring stations are scattered, it is very complicated to accurately predict the ozone level. Federated Learning (FL) allows everyone to train models together without exchanging raw data, thus protecting privacy. This paper uses FL to predict air quality, and the key point is to establish a prediction model, not to change the algorithm. This paper used UCI air quality data set (with 8,762 records, each with four attributes) to simulate the situation of five users, and the data were separated by independent identically distributed (IID) and non-independent identically distributed (Non-IID). This paper trained 50 communication rounds with a multi-layer perceptron (MLP) and Federal Average (FedAvg) method. The model of centralized training and local training is also used to compare. The experimental results show that the average absolute error (MAE) of FedAvg is 48.9 under IID data and 58.3 under Non-IID data, which is much better than the local training models (MAE 78.5 and 85.2) and close to the effect of centralized training (MAE 42.3 and 51.6). This method of federated learning adapts well to different data. This study confirms that FL is a feasible method for distributed air quality prediction.

Keywords: Federated Learning, Air Quality Prediction, FedAvg, Non-IID Data.

1. Introduction

Air pollution is a very common environmental problem. Ground ozone (O3) is an important secondary pollutant, which is produced by complex photochemical reactions of nitrogen oxides and volatile organic compounds. Accurate prediction of ozone level is very important for protecting public health and formulating environmental policies [1, 2].

Traditional air quality prediction methods are divided into two categories. One is mechanical model, which relies on detailed emission data and weather information [3, 4]. These models need a lot of computing resources, and it is difficult to update them continuously. The other is empirical model, which looks for rules from past observation data, but it is often difficult to deal with nonlinear and complex atmospheric data [5, 6]. Recently, the deep learning framework has been used for sequence prediction, and some progress has been made in the prediction accuracy [7, 8].

At present, most studies adopt the method of centralized training, that is, the data of all monitoring points are sent to a central station to make a unified model [9]. Although this method can make the best use of data, it has two main problems. First of all, different regions have different weather environments and pollution sources, which makes it difficult for a general model to take care of the special situation of each place [10]. Secondly, the process of collecting and sending data will bring troubles in privacy and data management, which often makes it impossible to exchange data across cities or organizations [11]. In order to solve these problems, some studies have trained models for each monitoring point separately [12]. This practice avoids the need to share data, but at the expense of those useful commonalities between different monitoring points. As a result, when the amount of data is small, these models are more prone to "over-fitting" problems, and their generalization ability is poor.

Federal learning provides a feasible solution to this problem. Its core principle is simple: the data is kept locally, and only the model parameters are sent to the central server for merging [13, 14]. This method not only protects the confidentiality of data, but also benefits from the knowledge of multiple data sources [15, 16].

According to this idea, this paper used federated learning to predict air quality in this study. The main purpose is not to change the method of federated learning, but to focus on the prediction model [17]. This method is chosen because in federated learning, the design and improvement of prediction model is particularly important for accuracy, but previous studies have not done much on this point. The work is mainly divided into three parts: (1) this paper sets up a multi-site federated learning system with real air quality data sets, and use FedAvg algorithm, and compare centralized training with single-site training to see whether federated learning is suitable for predicting air quality. (2) In order to see what effect the data distribution has, this paper has prepared two data partitions, IID and Non-IID, to make a comparison, focusing on how different data distributions affect the model performance. (3) This paper has done a series of comparative experiments to quantitatively measure the prediction accuracy and convergence speed of federated, centralized and local models, hoping that these findings will be helpful to practical application.

The rest of this article is arranged like this. Section 2 will talk about how the prediction model is designed and how this paper do comparative experiments. Section 3 will explain in detail where the data came from, how to process the data, how to set up the experiment, what results this paper found and how to understand them. Section 4 is a summary of the whole paper. At the same time, it will also talk about the limitations of the research and what direction this paper can continue to explore in the future.

2. Method

2.1. Dataset preparation

The data used in this analysis comes from UCI Air Quality Data Set, which was collected by an Italian air quality monitoring station from March 2004 to February 2005 [18]. The original data has 9,471 hourly records, including 15 different indicators, such as carbon monoxide (CO), temperature (T), relative humidity (RH) and absolute humidity (AH) [19]. The main target this paper wants to predict is the average ozone (O3) concentration per hour.

After deleting incomplete data, 8,762 complete records were finally left. According to the expert's suggestion, this paper selected four indicators as inputs: CO(GT), T, RH and AH, and then standardized all the selected variables by Z-score method [20]. Then the data are divided into training set (85%, 7,448 items) and test set (15%, 1,314 items).

To simulate the federated learning framework of five clients, the training data is divided into two different configurations. In IID configuration, data is randomly scrambled and evenly distributed, ensuring the same data distribution for each client [14]. In the Non-IID configuration, the data are sorted by ozone concentration and then distributed to clients in sequence, so that the data pollution degree of each client has a gradient from low to high, which is closer to the data difference in the real world [21].

2.2. FedAvg-based air quality prediction

The basic prediction model is a multi-layer perceptron (MLP). Its structure includes: an input layer (with 4 neurons), two hidden layers (with 64 and 32 neurons respectively, using ReLU activation function), and an output layer (with only one linear neuron) [17].

On the aggregation mechanism, the Federal Average (FEDAVG) is used [22]. The specific process is as follows: (1) The central server initializes a global MLP model; (2) Send this model to all five participating clients; (3) Each client trains the model with its own private data and runs several rounds of local iteration with SGD; (4) After the training, the client sends the updated model parameters back to the server;(5) the server performs weighted averaging:the server updates the global model by averaging the client parameters: $\mathrm{w}_{\mathrm{global}} = \frac{(\mathrm{w1}+\mathrm{w2}+\mathrm{w3}+\mathrm{w4}+\mathrm{w5})}{5}$; (6) steps 2-5 are repeated for 50 communication rounds.

2.3. Experimental hyperparameter configuration

This paper uses PyTorch to do the experiment. The setup is as follows: there are 50 exchange rounds, each round is trained locally for 5 times, and 32 samples are taken to learn together. The SGD optimizer is used, and the learning speed is 0.01 and the momentum is 0.9. MSE is used to measure the number of mistakes. To see if the effect is good, use MAE and RMSE as two indicators [23]. This paper compared three methods: one is centralized training with all the data together (this is the best case), the other is local training for each student (this is the worst case), and the other is federal learning with FedAvg.

3. Results and Discussion

3.1. Experimental results

Table 1 reports the MAE and RMSE of the three models under IID and Non-IID settings. The values are averaged over three random seeds.

Table 1. Test performance comparison (MAE and RMSE)

Model	IID MAE	IID RMSE	Non-IID MAE	Non-IID RMSE
Centralized	42.3	58.7	51.6	69.2
Local-only(average)	78.5	102.3	85.2	115.7
Federated (FedAvg)	48.9	65.4	58.3	78.6

Under the condition of independent and identically distributed (IID), the performance of federated model is almost as good as that of centralized method (48.9 vs. 42.3 MAE), while the performance of model trained only by local data is much worse (78.5 MAE). This shows that FedAvg can effectively integrate the information of evenly distributed data sets [22, 14].

When the data is non-IID, the performance of all models will be degraded. However, the federated model still has obvious advantages over the pure local model (58.3 vs. 85.2 MAE), which shows that it has stronger adaptability to data differences [21, 24]. This result is consistent with previous research: even without proximal terms, FedAvg can partially offset the influence of client drift [17].

3.2. Discussion

These findings tell three important points. First, federated learning is better than local training, because it expands the training pool by sharing parameters, which can reduce the problem of over-fitting [25]. Secondly, for FedAvg, there is little difference between IID and Non-IID (MAE increased from 48.9 to 58.3), which shows that a general model can also be useful in different settings [26, 27]. Thirdly, convergence analysis (omitted here for brevity) shows that the model will stabilize after 15 to 20 rounds, and this trend is the same as that seen in urban air quality prediction [24].

However, this study also has some limitations. The data used only comes from one place, so this paper not sure if it can be used in other areas [10]. This paper only tested the FedAvg method; More complex methods such as FedProx may have better results in the case of Non-IID [15]. The next work will add time mode (such as LSTM) and verify it on more data sets [28].

4. Conclusion

This study developed a federal learning system to predict air quality, using UCI Air Quality Dataset. This paper uses a method called FedAvg to optimize the neural network, try it on the same type and different types of data partitions, and compare it with the results of centralized training and single client training. The results show that the federated learning is obviously better than the local training method, and even if the data is unevenly distributed, it can get the same effect as the centralized model. The average absolute error of federated learning is 48.9 on the same type of data and 58.3 on different types of data, which shows that it is really powerful. However, there is a major limitation of this study, that is, the data set only comes from one region; Future research will use more city data, try more powerful methods such as FedProx, and add time series models such as LSTM to better understand the changing law of data.

References

[1] Jumin E, Zaini N, Ahmed AN, Abdullah S, Ismail M, Sherif M, et al. Machine learning versus linear regression modelling approach for accurate ozone concentrations prediction. Eng Appl Comput Fluid Mech. 2020;14(1):713–725.

[2] Mao W, Wang W, Jiao L, Zhao S, Liu A. Modeling air quality prediction using a deep learning approach: Method optimization and evaluation. Sustain Cities Soc. 2021;65:102567.

[3] Navares R, Aznarte JL. Predicting air quality with deep learning LSTM: Towards comprehensive models. Ecol Inform. 2020;55:101019.

[4] Zhang Z, Zhang S, Chen C, Yuan J. A systematic survey of air quality prediction based on deep learning. Alex Eng J. 2024;93:128–141.

[5] Du S, Li T, Yang Y, Horng SJ. Deep air quality forecasting using hybrid deep learning framework. IEEE Trans Knowl Data Eng. 2021;33(6):2412–2424.

[6] Carbajal-Hernández JJ, Sánchez-Fernández LP, Carrasco-Ochoa JA, Martínez-Trinidad JF. Assessment and prediction of air quality using fuzzy logic and autoregressive models. Atmos Environ. 2012;60:37–50.

[7] Zheng Y, et al. Forecasting fine-grained air quality based on big data. In: Proc 21st ACM SIGKDD Int Conf Knowl Discov Data Min; 2015. p. 2267–2276.

[8] Jin N, Zeng Y, Yan K, Ji Z. Multivariate air quality forecasting with nested long short-term memory neural network. IEEE Trans Ind Inform. 2021;17(12):8514–8522.

[9] Yi X, Zhang J, Wang Z, Li T, Zheng Y. Deep distributed fusion network for air quality prediction. In: Proc 24th ACM SIGKDD Int Conf Knowl Discov Data Min; 2018. p. 965–973.

[10] Bian C, Huang G. Federated Bayesian network approach for cross-regional air pollution classification: a case study of the Beijing-Tianjin-Hebei region. Environ Monit Assess. 2024;196(7).

[11] Putra KT, et al. Federated compressed learning edge computing framework with ensuring data privacy for PM2.5 prediction in smart city sensing applications. Sensors. 2021;21(13):4586.

[12] Kok I, Simsek MU, Ozdemir S. A deep learning model for air quality prediction in smart cities. In: IEEE Int Conf Big Data; 2017. p. 1983–1990.

[13] Chhikara P, Tekchandani R, Kumar N, Guizani M, Hassan MM. Federated learning and autonomous UAVs for hazardous zone detection and AQI prediction in IoT environment. IEEE Internet Things J. 2021;8(20):15456–15467.

[14] Liu Y, Nie J, Li X, Ahmed SH, Lim WYB, Miao C. Federated learning in the sky: aerial-ground air quality sensing framework with UAV swarms. IEEE Internet Things J. 2021;8(12):9827–9837.

[15] Roy S, Kandhoul N, Dhurandher SK. Generative adversarial network-based imputation and personalized federated learning with CNN-LSTM for AQI prediction. Eng Appl Artif Intell. 2025;158.

[16] Yang K, Jiang T, Shi Y, Ding Z. Federated learning via over-the-air computation. IEEE Trans Wirel Commun. 2020;19(3):2022–2035.

[17] Le DD, et al. Insights into multi-model federated learning: an advanced approach for AQI forecasting. Algorithms. 2022;15(11):434.

[18] Li X, Peng L, Hu Y, Shao J, Chi T. Deep learning architecture for air quality predictions. Environ Sci Pollut Res. 2016;23(22):22408–22417.

[19] Wardana INK, Gardner JW, Fahmy SA. Collaborative learning at the edge for air pollution prediction. IEEE Trans Instrum Meas. 2024;73:1–12.

[20] Gilik A, Ogrenci AS, Ozmen A. Air quality prediction using CNN+LSTM-based hybrid deep learning architecture. Environ Sci Pollut Res. 2021;29(8):11920–11938.

[21] Dey S, Raina R, Mishra S, Prasad AS, Dharavath R. GreenEdge AI: sustainable federated learning for smart city air quality prediction. J Ind Inf Integr. 2026;50:101081.

[22] Chhikara P, Tekchandani R, Kumar N, Tanwar S, Rodrigues JJPC. Federated learning for AQI prediction using UAV swarm networks. In: IEEE GLOBECOM; 2021.

[23] Feng Y. Participant privacy protection and air quality prediction based on FCM, PFI, LSTM, and DBN. Results Eng. 2025;26:105496.

[24] Dey S, Pal S. Federated learning-based air quality prediction for smart cities using BGRU model. In: ACM MobiCom; 2022.

[25] Beriwal S, Ayeelyan J. Decoding pollution: a federated learning-based pollution prediction study with health ramifications using causal inferences. Electronics. 2025;14(2):350.

[26] Huang W, Wang F, Chen J, Liu J, Liu Z, Li T. A federated framework for air quality prediction with predefined graph and adaptive graph. IEEE Internet Things J. 2025;12(14):27606–27617.

[27] Song M, Zhao X, Pedrycz W. A robust federated biased learning algorithm for time series forecasting. Cluster Comput. 2025;28(4).

[28] Sery T, Shlezinger N, Cohen K, Eldar YC. Over-the-air federated learning from heterogeneous data. IEEE Trans Signal Process. 2021;69:3796–3811.