# Comparison of Unilateral Algorithms Based on Federated Learning in Smart Cities

Zutao Zhang(✉), Junhong Lai, Fangze Cao, Yutong Guan, and Qian Zhu

Northeastern University, Shenyang 110819, China
20206858@stu.neu.edu.cn

**Abstract.** With the continuous popularization of the concept of smart cities, the environmental issues in smart cities have also received extensive attention. An important indicator reflecting the environmental problems in smart cities is the concentration of PM2.5 in the air. At the same time, we use a federated learning framework and try to select the appropriate algorithm in the federated learning edge learning device of the scene. For this reason, we compare decision tree algorithm, Gaussian regression algorithm, support vector machine algorithm, neural network algorithm, etc. Finally, through modeling and experiments, it is concluded that using Gaussian regression algorithm on edge devices is more effective.

**Keywords:** Federal learning · Machine learning · Smart city · Air quality

## 1 Introduction

With the rapid economic development and economic transformation and upgrading, the new concept of "smart city" has gradually become an important theoretical basis for the current environmental degradation, resource waste, traffic congestion and other issues. Under the concept of smart city, future urban development will inevitably be integrated with technology.

This paper mainly conducts related research by combining air quality, smart city and federated learning. Air pollution is a major challenge globally, especially in smart cities. Air pollution usually refers to the entry of certain substances into the atmosphere due to human activities or natural processes, showing sufficient concentrations for a sufficient period of time, and thus endangering human comfort and health, welfare and environmental phenomena [1]. PM2.5 issues at this stage can be observed in smart cities. Many smart cities face air pollution problems.

This paper presents a federated learning-based study of air quality prediction in smart cities. This study attempts to establish a new federated learning algorithm for horizontally dividing the dataset based on the data set of smart city air quality and the "client-server" architecture which based on federated learning. The established federated learning algorithm is evaluated by comparing the regression learning indicators calculated by previous machine learning and deep learning. In addition, this research aims to design and optimize the evaluation model for the federated learning client on the basis of federated learning, obtain the optimal solution of the data within a certain period of time, and verify its authenticity and effectiveness through simulation experiments.


D. Kumar et al. (Eds.): IEIT 2023, AHSSEH 10, pp. 106–111, 2023.
https://doi.org/10.2991/978-94-6463-230-9_13

## 2 Related Work

Before conducting the research, we first conducted a serious investigation on the previous work of federated learning in the field of smart cities.

In reviewing the research on federated learning methods in smart cities, Zhaohua Zheng [4] (2021) summarizes the application of federated learning in smart cities: recent progress, classification, possible challenges, connection science. Manzoor Ahmed Khan [6] (2021) discusses the potential interaction of different technologies and artificial intelligence to enable the properties required for future smart city services. After discussing some use cases for future eHealth, the design goals and technical requirements of the enabling technology are listed. The current situation of foreign logistics development is explained in depth. Ji Chu Jiang, Burak Kantarci [2] (2020) provide an overview of smart city sensing and its current challenges, and the potential of federated learning in addressing them. Dong Li [11] (2021) systematically studies privacy and security issues in the field of blockchain-based FL methods to provide an objective roadmap of the current state of the problem.

In the above research on the overview and summary of federated learning in smart cities and other aspects, these authors separately from the various methods of federated learning, the concept of smart city and various applications, and federated learning in IoT security and blockchain, etc. aspects are discussed and summarized. And in terms of digital twin technology for newer smart cities, Swarna Priya Ramu [3] (2022) discussed the concept, future development and future direction of federated learning in advancing digital twin technology in smart cities and introduced a method to integrate FL with DT (Digital Twin) to simplify the governance of smart cities, discussing the use of this integration in real-time and life-critical applications of smart cities still exists challenges and their possible solutions.

In the commercial and medical aspects of smart cities, federated learning has performed well. Weiqing Li [10] (2021) proposed a secure joint recommendation system scheme based on local differential privacy and secure aggregation to balance data privacy protection and training efficiency. At the same time, the performance of centralized training is achieved, and the effect of the system scheme is demonstrated. Shama Siddiqui, Anwar Ahmed Khan [9] (2020) discuss the application of federated learning in smart city modules to address resident body obesity and formulate rehabilitation plans.

More representative is the contribution of federated learning to edge computing and optimization. Basheer Qolomany [7] (2020) proposed a particle swarm optimization (PSO)-based technique to optimize the hyperparameter settings of local ML models in the FL environment, and used two case studies to evaluate the performance of their proposed technique. Karisma Trinanda Putra [5] (2021) proposed a novel edge computing framework that ensures data privacy for PM2.5 prediction in smart city sensing applications while providing efficient data generation. The scheme inherits the basic idea of compression technology and regional joint learning, and considers secure data exchange. Therefore, it can reduce the amount of data while protecting data privacy. Safa Otoum [8] (2021) proposes an adaptive framework that integrates federated learning and blockchain for network trustworthiness and security, probabilistically dealing with individual trust, and where security standards are met Estimate the trust value of end devices belonging to different networks. The predecessors mainly analyzed the application of federated learning in smart cities, combined with various methods of federated learning, the concept of smart cities and various applications. And also discussed and summarized the application of federated learning in respects of IoT security, smart medical care, blockchain and edge computing.

It can be seen that air quality in smart cities has always been a major challenge for cities. Therefore, intelligent air quality prediction systems will become an indispensable and important part of smart cities. The proposed intelligent air quality prediction system aims to effectively predict and analyze the air quality of smart cities, so as to achieve the purpose of subsequently improving the living environment and happiness index of residents. With the increase in the number of respiratory-related chronic patients worldwide, intelligent air quality prediction systems will become an important part of the future of smart cities.

## 3 Data Analysis

Introduction to data set: This data set includes EPA daily PM2.5 concentration data from monitoring stations where FRM and FEM are both in operation. This data set covers all states from 2016 to 2020. It also includes information about specific locations and local meteorological conditions. Monitoring and concentration data are extracted from EPA's air quality system (AQS) through API. The two key data sets downloaded are "Daily Summary" and "Monitoring" data.

In the process of data processing, we can easily find the data of a and b in the data column. We can find that they represent the comparison of the two instruments that detect PM2.5. Class a represents FRM and class b represents FEM (Fig. 1).

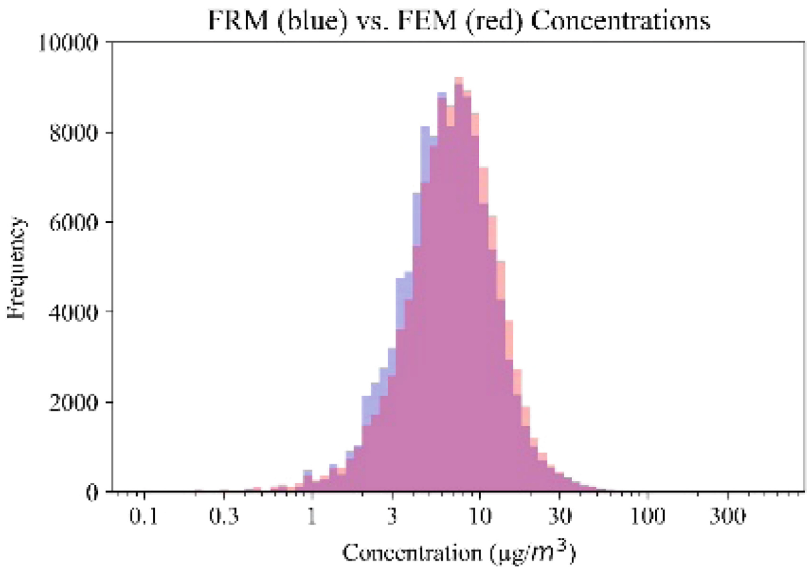


**Fig. 1.** FRM vs FEM

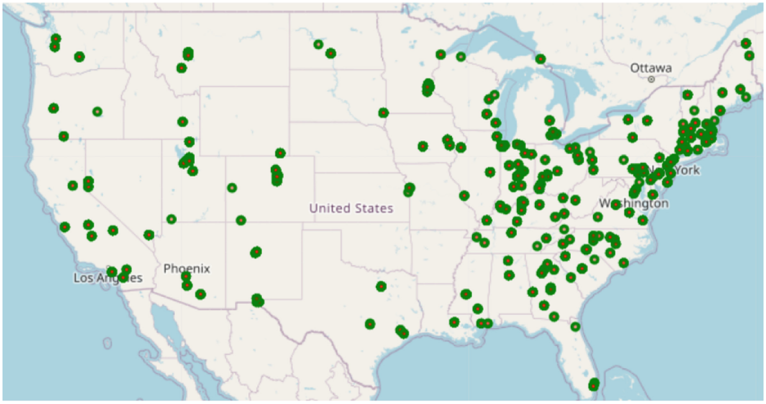


**Fig. 2.** Data distribution

We can find that the concentration of FRM is lft compared with that of FEM. It is not difficult to see that the predicted concentration of PM2.5 of FRM is lower than that of FEM on average.

Use the longitude and latitude of the dataset to visualize the dataset. Here, use the folium library in python to visualize the dataset. The effect is as follows (Fig. 2):

In general, we find that the data set is widely distributed and its distribution is relatively scattered. Among them, there are few sampling points in the Midwest of the United States, and many in the southeast of the United States, especially in the eastern coastal areas.

Therefore, we consider grouping by using the longitude and latitude in the dataset. There are 116712 pieces of data in the data set, and 266 groups of data after grouping by longitude and latitude.

Each group with a certain number can be used as an edge device for federated learning. By training the local model and uploading the model, the accuracy of the overall federated learning can be improved.

## 4 Model Building and Simulation Experiment

We use all the data to train and try to find a model that can accurately predict PM2.5 content. This model can be used as a reasonable and accurate prediction model on the edge device of federated learning in this scenario, which can be used to improve the accuracy of the overall federated learning.

As shown in the Table 1, the RMSE and MSE of the Gaussian regression model are lower than those of other models, so it can be basically considered that the Gaussian regression model performs best in this scenario.

**Table 1.** Model comparison

| | RMSE | MSE | MAE | Train Time |
|---|---|---|---|---|
| Robust Linear Regression | 5.5071 | 30.328 | 3.133 | 935.96 |
| Fine Tree | 6.4493 | 41.594 | 3.3058 | 200.43 |
| Rough Tree | 6.2625 | 39.219 | 3.2308 | 47.487 |
| Medium Gaussian SVM | 5.0854 | 25.861 | 2.7848 | 50219 |
| Rough Gaussian SVM | 5.3465 | 28.585 | 2.9523 | 51258 |
| Integrated Lifting Tree | 6.301 | 39.702 | 3.2345 | 758.56 |
| Integrated Bagging Tree | 5.9107 | 34.936 | 2.9712 | 2100.9 |
| GPR | 4.9307 | 24.311 | 2.9264 | 103220 |
| Narrow Neural Network | 7.4337 | 55.26 | 4.2018 | 53716 |
| Medium Neural Network | 7.8366 | 61.412 | 4.3915 | 54355 |
| Wide Neural Network | 7.5464 | 59.947 | 4.3184 | 57684 |
| Two-Layer Neural Network | 7.1189 | 50.679 | 3.5962 | 53885 |
| Three-Layer Neural Network | 6.8212 | 46.529 | 3.5992 | 54277 |

## 5 Conclusions

This article is aimed at the scenario of air quality prediction in smart cities. Because of the need to ensure the security of its data, the server-client architecture of federated learning is introduced. Under this framework, we focus on what algorithm can be used on the edge device of federated learning to achieve better results. Therefore, we compared the decision tree algorithm, support vector machine algorithm, Gaussian regression algorithm, and neural network algorithm on the data set, and finally found that Gaussian regression algorithm performs best for future generations to study.

**Acknowledgment.** This work has received funding from National Natural Science Foundation of China (NSFC) under grant agreement no.61902057.

## References

1. B. R. Subrahmanyam, A. Gautam Singh and P. Tiwari, "Air Purification System for Street Level Air Pollution and Roadside Air Pollution," 2018 International Conference on Computing, Power and Communication Technologies (GUCON), Greater Noida, India, 2018, pp. 518–522, https://doi.org/10.1109/GUCON.2018.8674934.
2. Ji Chu Jiang, Burak Kantarci, Sema Oktug and Tolga Soyata.Federated Learning in Smart City Sensing: Challenges and Opportunities. Sensors 2020, 20, 6230; https://doi.org/10.3390/s20216230

3. Ramu S P, Boopalan P, Pham Q V, et al. Federated learning enabled digital twins for smart cities: Concepts, recent advances, and future directions[J]. Sustainable Cities and Society, 2022, 79: 103663.
4. Zhaohua Zheng, Yize Zhou, Yilong Sun, Zhang Wang, Boyi Liu & Keqiu Li (2021): Applications of federated learning in smart cities: recent advances, taxonomy, and open challenges, Connection Science, DOI: https://doi.org/10.1080/09540091.2021.1936455.
5. Putra, K.T.; Chen, H.-C, Prayitno; Ogiela, M.R.; Chou, C.-L and Weng, C.-E.; Shae, Z.-Y. Federated Compressed Learning Edge Computing Framework with Ensuring Data Privacy for PM2.5 Prediction in Smart City Sensing Applications. Sensors 2021, 21, 4586. https://doi.org/10.3390/s21134586
6. Khan, M.A.; Alkaabi, N. Rebirth of Distributed AI—A Review of eHealth Research. Sensors 2021, 21, 4999. https://doi.org/https://doi.org/10.3390/s21154999.
7. Qolomany B, Ahmad K, Al-Fuqaha A, et al. Particle swarm optimized federated learning for industrial IoT and smart city services[C]//GLOBECOM 2020–2020 IEEE Global Communications Conference. IEEE, 2020: 1–6.
8. S. Otoum, I. A. Ridhawi and H. Mouftah, "Securing Critical IoT Infrastructures With Blockchain-Supported Federated Learning," in IEEE Internet of Things Journal, vol. 9, no. 4, pp. 2592–2601, 15 Feb.15, 2022, https://doi.org/10.1109/JIOT.2021.3088056.
9. Siddiqui S, Khan A A, Dev K, et al. Integrating federated learning with IoMT for managing obesity in smart city[C]//Proceedings of the 1st Workshop on Artificial Intelligence and Blockchain Technologies for Smart Cities with 6G. 2021: 7–12. https://doi.org/10.1145/3477084.3484950 .
10. W. Li, H. Chen, R. Zhao and C. Hu, "A Federated Recommendation System Based on Local Differential Privacy Clustering," 2021 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/IOP/SCI), Atlanta, GA, USA, 2021, pp. 364-369, doi: https://doi.org/10.1109/SWC50871.2021.00056.
11. Dong Li1, Zai Luo and Bo Cao. Blockchain-based federated learning methodologies in smart Environments. Cluster Computing, 2021, https://doi.org/10.1007/s10586-021-03424-y