# Efficiency-improved Federated Learning Approaches for Time of Arrival Estimation

Zeteng Huang
*School of Intelligent Systems Engineering*
*Sun Yat-Sen University*
Guangzhou, China
huangzt5@mail2.sysu.edu.cn

Meng Zhou
*School of Intelligent Systems Engineering*
*Sun Yat-Sen University*
Guangzhou, China
zhoum89@mail.sysu.edu.cn

Sheng Liu
*School of Intelligent Systems Engineering*
*Sun Yat-Sen University*
Guangzhou, China
liush235@mail2.sysu.edu.cn

Linlin You
*School of Intelligent Systems Engineering*
*Sun Yat-Sen University*
Guangzhou, China
youllin@mail.sysu.edu.cn

***Abstract*—Driven by artificial intelligence and edge computing, federated learning methods have become increasingly advanced and applied in various fields. Specifically, federated learning has been applied in various scenarios in intelligent transportation system (ITS) in which there are privacy issues. Travel time prediction is an important research problem in ITS. In this study, we apply wide-deep-recurrent models in a federated learning framework to predict estimated time of arrival (ETA) of vehicles, which could potentially improve the drivers and passengers' experience in travel service. We also apply periodic layers updating and representational consistent enhancing strategies in synchronous and asynchronous federated learning to improve the model efficiency. In addition, we propose a synchronous learning rate decay method to enhance the efficiency of federated learning. We integrate federated learning frameworks with wide-deep-recurrent learning and use a ride-sharing order records data with multi-dimensional information to predict travel time in a decentralized way to protect personal data privacy. Comparing with the selected baselines, FedAvgPR and FedAsyncPR models can reduce communication cost by about 14.22% and 44.41% to reach a target level of accuracy measured by mean absolute percentage error (MAPE). Moreover, synchronous learning-rate decay method not only can reduce the communication cost, but also improve the optimal prediction accuracy by about 1.80% in synchronous federated learning models. Overall, the optimal accuracies of the proposed models range from 8.42% to 10. 21%. The proposed federated learning models could be helpful for improving drivers travel experience while tackling the potential data privacy concerns.***

***Keywords*—estimated time of arrival, travel service, federated learning, learning rate decay method.**

## I. INTRODUCTION

With the booming development of artificial intelligence, edge computing, and other advanced technologies, Internet of Things (IoT) devices, mobile devices, robotic vehicles and roadside sensors have become widely ubiquitous and grown rapidly in numbers. These devices generate massive and valuable data during their daily communication, which are commonly gathered at a data center. This configuration, however, raises great challenge to the effective use of the distributed data and data privacy issues. In this context, federated learning has been proposed to provide secure modeling pipelines with little to no data sharing that leads to a highly efficient privacy-preserving solution [1]. At the same time, intelligent transportation system (ITS) has placed more emphasis on the system operations instead of only focusing on infrastructure development [2]. And the analysis and application of large-scale and real-time traffic data is playing a vital role in ITS.

Estimated time of arrival (ETA) refers to the expected travel time between a pair of origin and destination along a given route, and its prediction is an important research topic in ITS. The accurate and fast travel time prediction can improve the operation performance of advanced traffic management system and the travel experience of travelers. Many machine learning methods have been applied for accurate prediction of ETA in various transportation domains. For instance, reference [3] introduced a system which learned from historical trajectories and used the 3D grid points to collect key feature to estimate flights travel time through regression models and recurrent neural network (RNN). A plug and play deep learning based generative model was utilized to update the ETA information of the buses on the go [4]. Taking advantage of heterogeneous information graph, reference [5] translated the road map into a multi-relational network and introduced a trajectory based network to jointly consider the travel behavior pattern for the ETA prediction of vehicles. To address the data sparsity problem, the road network metric learning framework consisted of a main regression task to predict the travel time and an auxiliary metric learning task to improve the quality of link embedding vectors was proposed [6]. Reference [7] presented a tree-structured Long Short-Term Memory (LSTM) model with attention mechanism to predict vehicles' travel time. Besides, the proposed model substitutes a tree structure with attention mechanism for the unfold way of standard Long Short-Term Memory to construct the depth of LSTM and modeling long-term dependence. The studies mentioned above are all training their models in a centralized manner so that all data need to be gathered in data center. As such, the risk and damage of data leakage are high and the performance of the model is limited by the level of concentration of data. To tackle the issue of data availability and security, a decentralized training mode, federated learning (FL), has received increasing popularity and has been widely applied to various task in ITS. An original FL framework, FASTGNN, was proposed for preserving transportation networks' topological information while protecting the data privacy, and it was applied to forecast vehicles' speed [8]. Reference [9] put forward an FL based gated recurrent unit neural network algorithm with a trade-off between accurate traffic flow prediction and preserving data privacy.

Up to now and to the best of our knowledge, there is no research on predicting the estimated time of arrival (ETA) of vehicle trips in the mobility service context based on federated learning. However, the drivers' driving information is dispersed and highly private. The centralized learning approach may be hindered by data security and data privacy problems. On the contrary, federated learning, a decentralized machine learning framework, is suitable to leverage the driving information without compromising privacy. In this paper, we will apply a novel wide-deep-

2023 8th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS) | 978-1-6654-5530-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/MT-ITS56129.2023.10241670

recurrent (WDR) [10] learning model within a federated learning architecture to estimate travel time and propose a synchronous learning rate decay method to improve model efficiency and accuracy in synchronous and asynchronous federated learning models.

## II. Methodology

### *A. Wide-Deep-Recurrent Learning*

The central task in this study is to predict the travel time and, in turn, the time of arrival based on available information provided from trip records. Similar to many other cases, various different types of information could be extracted from individual-level trip records, including spatial information, temporal information, traffic information, personalized information and augmented information. For instance, spatial information includes traffic light, intersection and road segment records. Departure time, as temporal information, indicates when the trip started. Traffic information usually contains the traffic condition through the congestion classification. Personalized information depicts a specific driver who may have different driving preference. Other auxiliary information such as weather may also be provided. These features are in different shapes and sizes and can be respectively divided into dense features, sparse features and sequential features accordingly.

Because of the wide variety of features, a WDR learning model structure is well-suited for the prediction tasks. A WDR model contains wide, deep and recurrent models and is selected as the base model for the prediction of ETA. The model structure is as shown in Fig. 1. This model combines the advantages of its sub-models and can effectively utilize the dense features, the sparse features with high dimensions and the sequential features along the road segments. The WDR model in this study is adopted from [10] and consists of three main blocks:

- Wide model includes cross-product transformations followed by affine transformations with ReLU activations to produce a 256-dimensional effect.
- Deep model embeds sparse features into 16 dimensions and sets up a 3-hidden-layer MLP with ReLU activation to process them into a 256-dimensional output.
- The recurrent model is a 2-layer Long Short-Term Memory (LSTM) neural network. Firstly, each road segment is projected into a 256-dimensional space with a fully connected layer. Then, its transformed sequential characteristics are fed into LSTM with 256 units. Lastly, LSTM produces a final hidden state with 256 dimensions as its output.

Finally, the models mentioned above will be concatenated to estimate travel time through a fully connected layer.

In WDR, measuring accuracy of the model is through mean absolute percentage error (MAPE) loss which is the benchmark of back-propagation model parameters. And the target is to minimize the MAPE as formula (1):

$$\min_f \sum_{i=1}^{N} \frac{|y_i - f(x_i)|}{y_i} \tag{1}$$

where $y_i$ is actual travel time of the route $x_i$, $f(x_i)$ is the ETA of the route $x_i$, and $f$ is the overall regression model.

### *B. Federated Learning Framework for WDR*

In this section, we introduce the several proposed federated learning frameworks designed for WDR model. In essence, the federated learning framework consists of local clients that generate and maintain the data, a training model for the central learning task that is shared across clients, and a cloud server in charge of organizing the modeling process. Local clients, the training model and the global cloud server are participant drivers, the WDR model and the global federated aggregating algorithms in this study, respectively. Several federated learning frameworks are described as below:

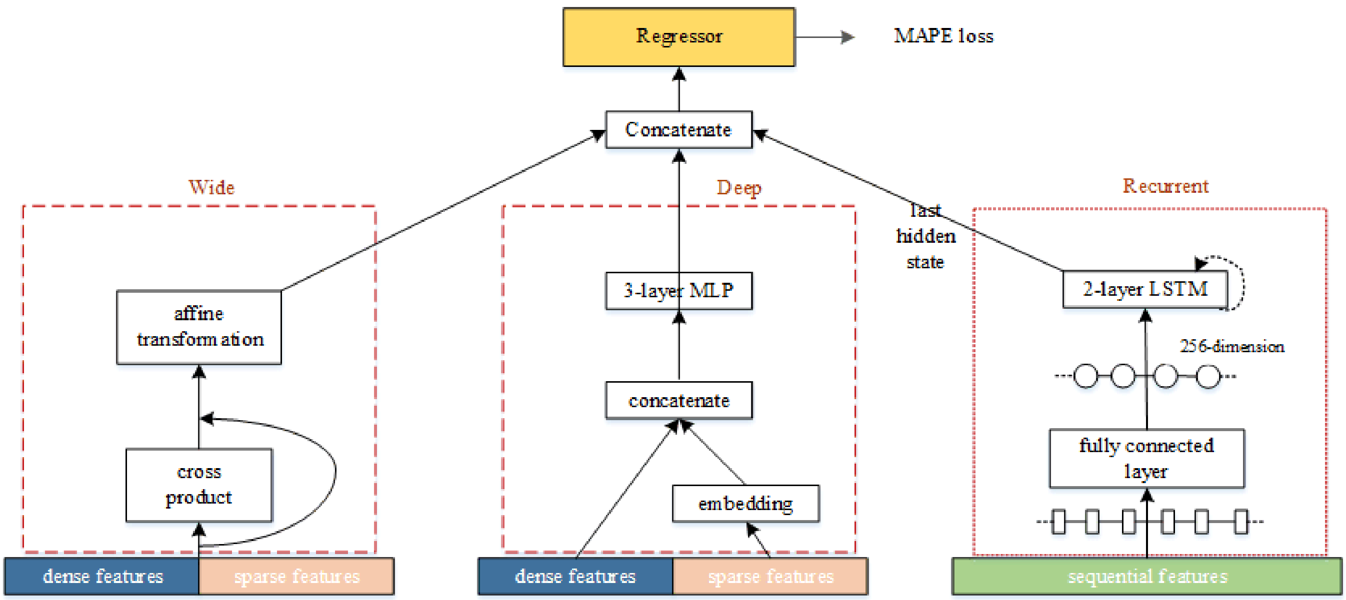



Fig. 1. The wide-deep-recurrent learning framework.

- FedAvg [11] is a widely adopted synchronous federated learning framework. It randomly selects a portion of clients which participate in a global round $r$ to upload and aggregate local models weights through weighted average method. And it is one of baselines in our experiments.

- FedAsync [12] is a classic asynchronous federated learning framework. It incorporates a proximal term $\mu$ that helps to amend the loss resulting from the hysteretic clients. And a mixed hyper-parameter $\alpha$ is used for balancing the current model weights and global model weights when the cloud server is updating the global model. In this paper, $\alpha$ is set to 0.5 [12] and $\mu$ is set to 0 as generated by the differences between local and global model weights. This is the other one of baselines of our experiments.

- FedAvgPR is a novel federated learning framework. It combines FedAvg with periodic layers uploading (PLU) strategy and representational consistency enhancing (RCE) strategy [13] [14]. In PLU, every global round is a training period, and the layers of deep neural networks will only be uploaded in specific global rounds. So, it can decrease the cost of communication. RCE, on the other hand, measures each layer of deep neural networks of local model's importance in a global round to determinate whether the received layer should be uploaded or not. Note that RCE is computed by the Pearson correlation coefficient [15] of local model layers' weights and global model layers' weights.

- FedAsyncPR is also a novel federated learning framework, in which the PLU and RCE strategies are incorporated with the FedAsync framework.

As illustrated in Fig. 2, the selected clients-set k, according to the rule-of-thumb [16], generates local data, and we select m clients as the training and integrated testing dataset.

As shown in Fig.2.(a), the global server firstly selects m clients from k clients randomly as the participants at a global round, according to the synchronous federated learning scheme. Then, the model in every selected client is trained locally and local model weights are uploaded. After that, the cloud server will aggregate k clients' weighted averaging local model weights. Lastly, the cloud server will send updated global model to local clients for the next round's training. Specifically, FedAvgPR model uploads and aggregates weighted averaging local model layers weights with bigger RCE value.

As shown in Fig.2.(b), on the other hand, the asynchronous federated learning framework selects m clients from k clients, ordered by time as the participants in a global training round. Every selected client trains its model and upload its model weights. The cloud server then aggregates composite local and global model weights according to a predefined rule from k clients. Lastly, the cloud server dispatches the updated global model to participant local clients in this global round. Specifically, FedAsyncPR model uploads and aggregates composite global and local model layers weights according to a predefined rule with bigger RCE value.

Through the above-mentioned decentralized learning process, the global model converges to the optimal condition without accessing the complete dataset. At the final stage, we can use combined testing dataset to measure the accuracy of the global federated learning model from each round.

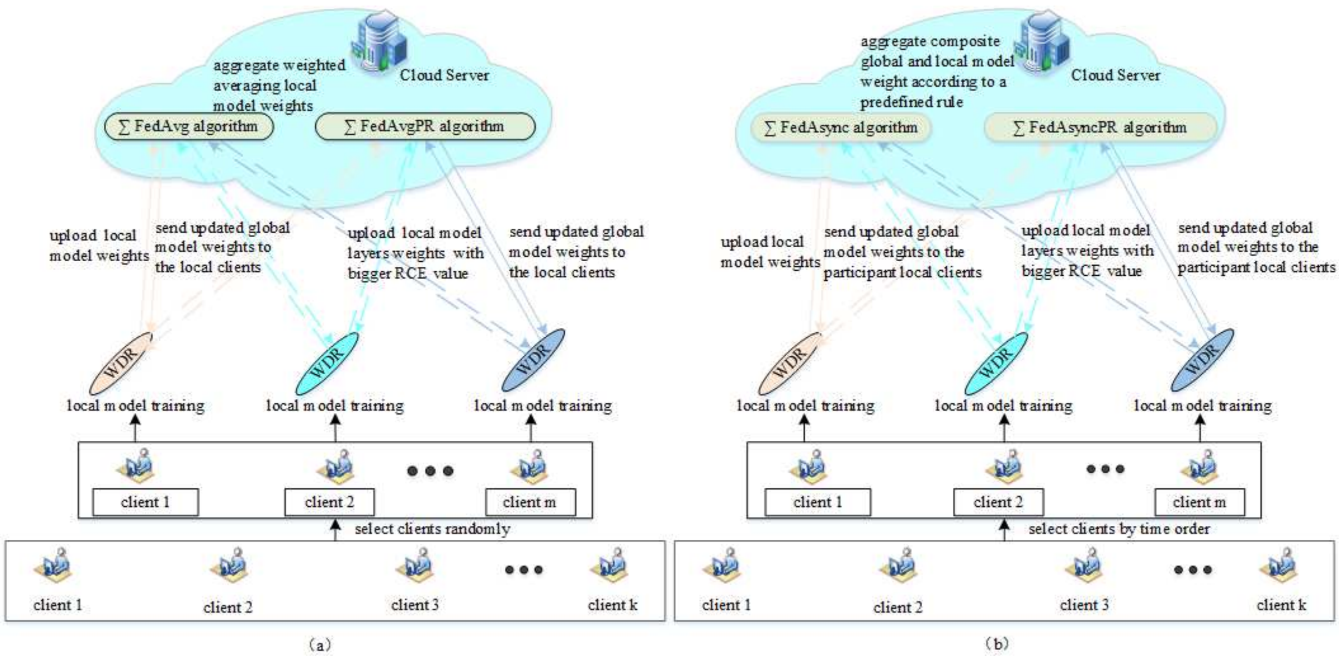



Fig. 2. (a) The synchronous federated learning framework for WDR; (b) The asynchronous federated learning framework for WDR.

### C. Learning Rate Decay Methods for Federated Learning

This study also proposes a learning rate decay approach for federated learning. Gradient descent families and adaptive optimization algorithms are the mainstream optimizers in machine learning. Among them, Adam [17] is one of the most popular adaptive optimizers and is selected as the learning rate decay baseline in this paper. Adam's adaptive learning rate decay is calculated as formula (2):

$$\eta_{ad} = \frac{\eta}{\sqrt{\hat{v}_t + \epsilon}} \hat{m}_t \tag{2}$$

where $\eta_{ad}$ denotes the learning rate after decaying, $\eta$ denotes initial learning rate, $\epsilon$ represents a small enough constant to guarantee a non-zero denominator, and $\hat{m}_t$ and $\hat{v}_t$ are the amended first and second moment estimation.

During the experiment, we found that a relatively larger learning rate is needed for convergence. But the vanilla Adam's adaptive learning rate decay method is prone to low convergence or even divergence. In this case, we propose an additional synchronous learning rate decay method besides Adam optimizer to improve the accuracy and communication cost reduction of the models. And it can be expressed as formula (3):

$$\eta_s = \alpha^{\frac{r}{10}} \eta \tag{3}$$

where $\alpha$ denotes the coefficient for determining the rate of decay.

In this case, the modified Adam's adaptive learning rate should be calculated as formula (4):

$$\eta_{ad} = \frac{\eta_s}{\sqrt{\hat{v}_t + \epsilon}} \hat{m}_t \tag{4}$$

## III. EVALUATION AND DISCUSSION

In this section, we will firstly introduce the model's basic settings, e.g., the dataset that we use, the model to be trained firstly. Secondly, we will evaluate the efficiency improvement of proposed FedAvgPR and FedAsyncPR using FedAvg and FedAsync as baselines. And the efficiency and accuracy enhancement of the proposed additional learning-rate decay methods will be evaluated in comparison with adaptive learning-rate decay in designed experiments. Finally, we will discuss the results of the implemented experiments.

### A. Basic Settings

The dataset used is the ride-sharing order records for the estimated time of arrival challenges from Didi Gaia open data scheme (https://gaia.didichuxing.com; accessed from SIGSPATIAL 2021 GISCUP). This floating car dataset is from Didi Chuxing, in Shenzhen, China in August, 2020. It includes departure time, traffic information, weather information and so on. We processed two weeks data and chose 100 drivers' order data according to the rule-of-thumb as the training and testing dataset and 20 participant clients in a global round. A brief introduction of the extracted data features including journey time, distance, sum link time, sum cross time, speed, driver id, distance classification, sum crosses, day of week, hour category, weather and link arrival status is listed in TABLE I. In this paper, we will implement 8 federated learning experiments and the key parameters are listed in TABLE II. And two indicators are used as the evaluation criterion, namely:

- $I_1$: The optimal MAPE value in 1000 global rounds is used to evaluate the experiments' accuracy performance.
- $I_2$: The number of global round when MAPE value reaches the predefined target MAPE $T_{mape}$. Note that $T_{mape}$ is set to 10.50%, because when MAPE is decreasing toward and around $T_{mape}$, the accuracy improvements of most models are rather low.

TABLE I. THE INTRODUCTION OF EXTRACTED DATA FEATURES

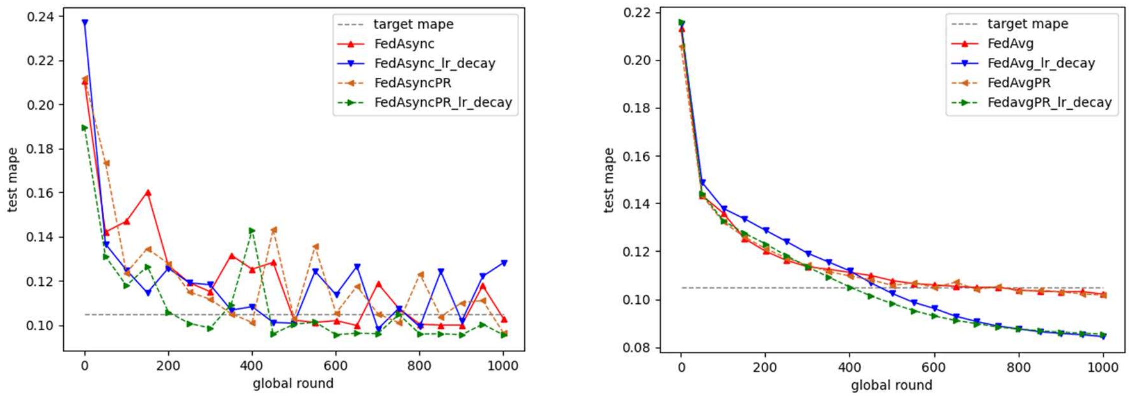
| Feature | Type | Description |
|---|---|---|
| journey time | dense | The actual trip time. |
| distance | dense | The actual trip distance. |
| sum link time | dense | The total time of passing every road segment. |
| sum cross time | dense | The total time of passing every signalized intersection. |
| speed | dense | The average speed. |
| driver id | sparse | The driver's id information. |
| distance classification | sparse | Trip distance classified into levels. |
| sum crosses | sparse | The total number of passed intersections. |
| day of week | sparse | The sorted day of a week. |
| hour category | sparse | 1 denotes rush and 0 denotes leisure. |
| weather | sparse | 0 denotes cloudy, 1 denotes moderate and 2 denotes shower. |
| link arrival status | sequential | The sequential numbers denote every road segment traffic condition. |

TABLE II. THE BASIC SETTINGS FOR EXPERIMENTS

| WDR Epochs | Global Rounds | Learning-rate | Models |
|---|---|---|---|
| 5 | 1000 | Adaptive learning-rate decay | FedAvg<br>FedAsync<br>FedAvgPR<br>FedAsyncPR |
| | | Adaptive and additional synchronous learning-rate decay | FedAvg<br>FedAsync<br>FedAvgPR<br>FedAsyncPR |

TABLE III. SUMMARY OF EXPERIMENTS RESULTS

| Learning-rate | Models | Optimal MAPE in 1000 Rounds | Round Reached $T_{mape}$ |
|---|---|---|---|
| Adaptive learning-rate decay | FedAvg | 10.21% | 668 |
| | FedAvgPR | 10.17% | 573 |
| | FedAsync | 9.57% | 394 |
| | FedAsyncPR | 9.52% | 219 |
| Adaptive and additional synchronous learning-rate decay | FedAvg | 8.42% | 472 |
| | FedAvgPR | 8.55% | 404 |
| | FedAsync | 9.51% | 321 |
| | FedAsyncPR | 9.47% | 202 |




(a) (b)

Fig. 3. (a) The asynchronous federated learning test MAPE, (b) The synchronous federated learning test MAPE.

## B. Evaluation of Federated Learning Experiments

The experiment results are summarized in TABLE III. It includes results using FedAvg, FedAvgPR, FedAsync and FedAsyncPR with the two kinds of learning-rate decay methods.

As summarized in TABLE III. the accuracies of the 8 models are in range from 8.42% to 10. 21%. The efficiency of FedAvgPR and FedAsyncPR is superior to FedAvg and FedAsync, respectively. FedAvgPR and FedAsyncPR can reach $T_{mape}$ at 573rd and 219th global round with efficiency improvements of 14.22% and 44.41% by adaptive learning-rate decay method. And they can reach $T_{mape}$ at 404th and 202nd global round with an efficiency improvements of 14.41% and 37.07% by a composite learning-rate decay method. On the other hand, the additional synchronous learning-rate decay method can also improve the models' efficiency. FedAvg and FedAvgPR with additional synchronous learning-rate decay can reach $T_{mape}$ at 472nd and 404th global round with an acceleration of 29.34% and 29.49%. In the meantime, FedAsync and FedAsyncPR can reach $T_{mape}$ at 321st and 202nd global round with an acceleration of 18.53% and 8.42%.

In addition, the optimal accuracy of asynchronous federated learning models is smaller than that of the synchronous federated learning models with adaptive learning-rate decay. The test accuracy is about 9.50% in asynchronous federated learning and 10.20% in synchronous federated learning. However, the optimal accuracy of asynchronous federated learning models is bigger than that of the synchronous federated learning models with adaptive learning-rate decay. The test accuracy is about 9.50% in asynchronous federated learning and 8.50% in synchronous federated learning. And it demonstrates that the asynchronous federated learning models can reach $T_{mape}$ at a much faster rate.

In Fig. 3 (a), the accuracy curve of asynchronous federated learning model is fluctuant. However, the range of accuracy curves by asynchronous federated learning models integrated with PLU and RCE strategies is smaller than that based on asynchronous federated learning models. Furthermore, the additional learning rate decay method can also reduce the fluctuation of the accuracy curve fluctuation. In Fig. 3 (b), it is obvious that the accuracy curves are decreasing smoothly. Not only are the decreasing rate of the synchronous federated learning models with additional learning-rate decay method higher, but the optimal accuracy values are smaller.

In summary, the above experiment results show that federated learning integrated with PLU and RCE and adding additional learning-rate decay method can reduce the communication cost significantly. PLU and RCE strategies can narrow the range of accuracy curve of asynchronous federated learning models and additional learning rate decay method can also improve the accuracy of synchronous federated learning models.

## C. Discussion

First, the results of the federated learning integrated with PLU and RCE indicate that the communication cost can be reduced through uploading local model weights, especially, the deep layers of neural networks in specific global rounds. Moreover, additional synchronous learning-rate method can help to reduce global round to reach $T_{mape}$, which could be leveraged to evaluate the communication cost reduction.

Second, the asynchronous federated learning models integrated with PLU and RCE strategies have lower levels of fluctuation in terms of accuracy. The reason is that the larger differences between local model weights and global model weights will lead to a smaller RCE value so that the global model will not aggregate the layers of relevant local model layers weights. Then the differences between local and global model weights may be narrowed.

Third, additional synchronous learning rate decay method can improve the accuracy of synchronous federated learning models. But it is not well-behaved to improve the model accuracy in asynchronous federated learning models, largely due to its fluctuation in modeling accuracy. The accuracy differences between adjacent models are usually large so that learning rate decay method is less helpful in improving the accuracy for asynchronous models.

Finally, it is obvious that communication cost reduction rate of PLU and RCE strategies is higher in asynchronous federated learning than in synchronous federated learning. The primary cause is the greater differences between local model weights and global model weights resulting from the asynchronous hysteresis which leads to smaller RCE value. Whether to upload local layers of model weights or not depends on RCE value and the predefined threshold value.

## IV. Conclusion

In this paper, we introduced federated learning frameworks integrated with wide-deep-recurrent learning method to estimate vehicle travel time in a privacy-preserving manner. In addition, we proposed a synchronous learning-rate decay method to improve the efficiency and accuracy of federated learning models.

Comparing with the two state-of-the-art baselines, the optimal accuracies of the proposed FedAvgPR and FedAsyncPR models have largely maintained at approximate levels. But their accuracy improvement is visibly faster. It also shows that the PLU and RCE strategies can immensely reduce the communication cost in synchronous federated learning models by about 14.00% and it is even more significant in asynchronous federated learning way by about 40.00%. The synchronous federated learning models with additional learning rate decay method can reduce the communication cost in asynchronous and synchronous federated learning models by about 30.00% and 13.00%, respectively. And it can also improve the accuracy of synchronous federated learning models by about 1.80%. Results suggest that the proposed federated learning models can be utilized to improve the travel experience of drivers and passengers in ride-sharing and other transport services effectively with personal data privacy preservation.

In the future, we plan to aim at improving the accuracy of the federated learning models while maintain the reduction in communication cost. In addition, we wish to implement the proposed federated learning models on larger-scale datasets and more client drivers for practical uses. Moreover, we also wish to test the generalization ability of the proposed federated learning models on different traffic datasets.

## Acknowledgment

This study was partially supported by the National Natural Science Foundation of China (42201495), the Guangdong Basic and Applied Basic Research Foundation (2021A1515110049), and the Fundamental Research Funds for the Central Universities, Sun Yat-Sen University (22qntd1703).

## References

[1] KhoKhar F A, Shah J H, Khan M A, et al. A review on federated learning towards image processing[J]. Computers and Electrical Engineering, 2022, 99: 107818.

[2] Lin Y, Wang P, Ma M. Intelligent transportation system (ITS): Concept, challenge and opportunity[C]//2017 ieee 3rd international conference on big data security on cloud (bigdatasecurity), ieee international conference on high performance and smart computing (hpsc), and ieee international conference on intelligent data and security (ids). IEEE, 2017: 167-172.

[3] Ayhan S, Costas P, Samet H. Predicting estimated time of arrival for commercial flights[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018: 33-42.

[4] Paliwal C, Biyani P. To each route its own ETA: A generative modeling framework for ETA prediction[C]//2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, 2019: 3076-3081.

[5] Hong H, Lin Y, Yang X, et al. HetETA: Heterogeneous information network embedding for estimating time of arrival[C]//Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining. 2020: 2444-2454.

[6] Sun Y, Fu K, Wang Z, et al. Road network metric learning for estimated time of arrival[C]//2020 25th International Conference On Pattern Recognition (ICPR). IEEE, 2021: 1820-1827.

[7] Ran X, Shan Z, Fang Y, et al. An LSTM-based method with attention mechanism for travel time prediction[J]. Sensors, 2019, 19(4): 861.

[8] Zhang C, Zhang S, James J Q, et al. FASTGNN: A topological information protected federated learning approach for traffic speed forecasting[J]. IEEE Transactions on Industrial Informatics, 2021, 17(12): 8464-8474.

[9] Liu Y, James J Q, Kang J, et al. Privacy-preserving traffic flow prediction: A federated learning approach[J]. IEEE Internet of Things Journal, 2020, 7(8): 7751-7763.

[10] Wang Z, Fu K, Ye J. Learning to estimate the travel time[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018: 858-866.

[11] McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Artificial intelligence and statistics. PMLR, 2017: 1273-1282.

[12] Xie C, Koyejo S, Gupta I. Asynchronous federated optimization[J]. arXiv preprint arXiv:1903.03934, 2019.

[13] Liu S, Chen Q, You L. Fed2A: Federated Learning Mechanism in Asynchronous and Adaptive Modes[J]. Electronics, 2022, 11(9): 1393.

[14] You L, Liu S, Chang Y, et al. A triple-step asynchronous federated learning mechanism for client activation, interaction optimization, and aggregation enhancement[J]. IEEE Internet of Things Journal, 2022.

[15] Benesty J, Chen J, Huang Y, et al. Pearson correlation coefficient[M]//Noise reduction in speech processing. Springer, Berlin, Heidelberg, 2009: 1-4.

[16] Alwosheel A, van Cranenburgh S, Chorus C G. Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis[J]. Journal of choice modelling, 2018, 28: 167-182.

[17] Wang W, Taylor J, Bala B. Exploiting the Power of Levenberg-Marquardt Optimizer with Anomaly Detection in Time Series[J]. arXiv preprint arXiv:2111.06060, 2021.