

Adaptive Federated Learning for Electric Power Inspection with UAV System

Yu Liang, Ruifan Huang, Xun Li, Junjie Yang, Xinkai Zhang, Xuehe Wang*

School of Artificial Intelligence, Sun Yat-sen University, Zhuhai, China

Abstract

With the rapid development of the national power grid, the demand for efficient and reliable power supply is increasing. As the labor cost and the size of the power grid are increasing, the Unmanned Aerial Vehicle (UAV) power inspection is a new and efficient way of detecting power grid abrasion. By analyzing the data collected by the UAVs, a smart detection and maintenance service can be provided. To improve the model robustness and accuracy, data collected from different companies and regions are required, which may violate the data privacy policy. As a distributed machine learning technique, Federated Learning (FL) can collaboratively train global models without sharing private data. In this article, in order to protect the data privacy between different systems, optimize models' accuracy and convergence performance for non-Independently-and-Identically-Distributed (non-IID) data, we propose an adaptive method that jointly adjusts the learning rate and gradient based on the idea of FL. By recording global gradient information and using the momentum to accelerate the training process, our method adaptively controls the local gradient and learning rate in training of local models, and be more robust to local minima. Finally, we verify the superiority of our model compared with the generic FL model for non-IID data through experiments.

Keywords:

Adaptive Federated Learning; Non-IID Data; UAV; Smart Grid.

1. Introduction

1.1. Motivation and Background

Recent times, China's power grid is undergoing rapid development. During the 12th Five-Year period, the scale of China's power grid has jumped to the first place in the world. So far, China has built six major cross-provincial power grids with a total transmission line length of over 1.15 million kilometers. However, the automation of distribution is in its infancy, and it takes a long time for fault diagnosis, isolation and recovery. Purely by manual work in such a long line inspection, patrol personnel will face large workload, high labor intensity, long patrol time, low patrol efficiency and other issues, and the complex structure of the power grid will make maintenance becomes a certain risk.

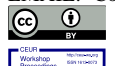
At present, the Unmanned Aerial Vehicle (UAV) cruise detection is mainly used to detect the power grid fault, return and repair the fault problem. That is, a new UAV-enabled inspection method of automatic, intelligent, high efficiency and supervision is introduced [1].

Distribution network involves different departments, and the data collected in different systems is inconsistent. In addition, there is a lack of data sharing mechanism and information acquisition channels. This is reflected in the lack of management refinement and the limited exchange of data, graphics and information.

To solve the problem, a kind of distributed machine learning technology called Federated Learning (FL) is introduced [2]. FL's core idea is to train the distributed models across multiple data sources, which offers the possibility to construct global models based on virtual fusion data by exchanging

ICCEIC2022@3rd International Conference on Computer Engineering and Intelligent Control

EMAIL: *Corresponding author: wangxuehe@mail.sysu.edu.cn (Xuehe Wang)



© 2022 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

model parameters and intermediate results without exposing the local source data to each other. FL approaches a unique way to balance the data sharing and data privacy protection, making data "available but not visible".

In previous distributed models, it is often assumed that the participants had Independently-and-Identically-Distributed (IDD) dataset. However, such an ideal scenario is often not available in most realistic scenarios. Participants are usually quite different from each other, as whose data are often non-IDD in practical questions. For example, in the power grid distributed across provinces, the climate in different regions causes different losses to cables. It also indicates that the server and participants need to communicate and exchange more times to achieve the required model accuracy. Unsatisfactory convergence performance, high communication cost and privacy guarantee also pose challenges to the optimization of FL's non-IDD training model.

1.2. Related work and Our idea

In the current research of FL, many training methods for models based on non-IDD data have been proposed. FedAVG [2] is the most classical and widely used federated optimization method, which can effectively reduce the communication cost compared with the traditional stochastic gradient descent (SGD) model. However, as it uses relatively static parameters, its convergence effect fluctuates greatly in the face of different optimization problems, thus it does not always have good enough convergence performance in the face of data with a higher heterogeneity [3].

In response to this problem, some adaptive federated optimization methods have attracted extensive attention. [4] proposed dynamic learning rate (DLR), which improved FedAVG algorithm by optimizing local learning rate to adapt the fading channel and realize efficient aggregation of wireless data. [5] proposed an adaptive data enhancement framework for imbalanced distributed training data to reduce communication traffic and accelerate convergence. [6] synthesized the current general ideas of adaptive optimization for FL and summarized the adaptive methods such as FedADAM, FedADAGRAD, FedYOGI by adaptively adjusting the learning rate.

All related works list above extend our research ideas, nevertheless, few articles discuss the direction of integrated adaptive optimization of learning rate and gradient. In the training of deep learning network, the gradient and learning rate are both factors of great significance. The convergence performance of the model will be greatly improved theoretically by adapting them in local training and global aggregation. More specific method will be presented in next section.

2. Adaptive Optimization

Federated Learning (FL) was approached by B. McMahan et al. in 2016 as a decentralized machine learning mechanism [2], which is trained jointly by a central server coordinating a set of distributed participating devices (which we refer to as clients). It avoids the direct aggregation of source data and protects the privacy of user data.

2.1. Generic Algorithm

FedAVG algorithm is a classical federated learning algorithm [2]. This algorithm proposes the basic idea of federated learning based on stochastic gradient descent algorithm (SGD).

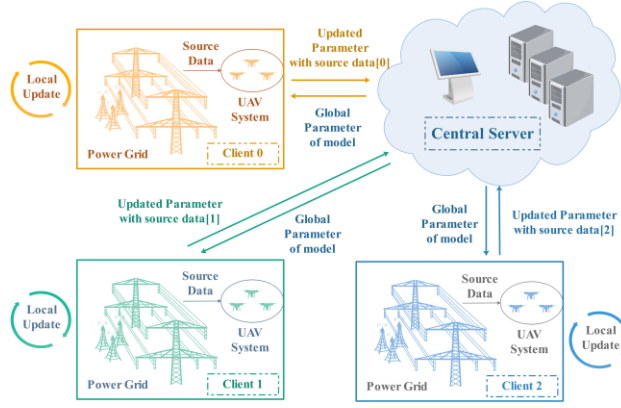


Figure 1 Background Architecture with FL

Assume that T is the total communication rounds, where $t = \{1, 2, \dots, T\}$, and the global parameter in the t -th round is $w(t)$. At the beginning of each communication round, K clients are randomly selected as participating samples, where $k \in \{1, 2, \dots, K\}$. Sample clients has its local dataset size D_k , then the total amount of sample data is denoted as $D = \sum_{k=1}^K D_k$.

In each communication round, the client uses its source data locally to perform one-step gradient descent on the current model to obtain the model parameters of the client:

$$w_k(t+1) = w(t) - \eta \nabla F_k(w(t))$$

where $\nabla F_k(w(t))$ is the gradient at iteration t of client k , η is the learning rate of the model. After receiving local parameters, servers perform model average operation to update global parameters:

$$w(t+1) = \sum_{k=1}^K \frac{D_k}{D} w_k(t+1)$$

The updated global parameters are synchronized to all clients for the next round of local training, and the process is repeated (Show in **Figure 1**).

Compared with the generic FedSGD model, the Fe-dAVG algorithm is much more accurate and robust to non-IID data to a certain extent. However, as it adopts static learning rate, gradient and other parameters in local training, its convergence speed will still be slow in the case of more imbalanced non-IID data in practical problems. There is still room for its improvement in this case [8].

Recent times, the idea of adaptively updating static parameters has spawned many optimization methods of FL, which can achieve faster convergence on non-IID datasets while ensuring robustness. FedADAM [6] is one of the most advanced algorithms. Based on SGD model, it adaptively adjusts the learning rate according to the momentum information of the gradient by tracking and calculating the first moment and second moment of gradient parameter of the model, and use Adam's update method for iteration of its global parameter [7]. Compared with standard SGD, these features enable FedADAM to converge faster and be more robust to local minima [6][9].

2.2. Our Adaptive Optimized Method

Communication cost plays a dominant role in the optimization of federated learning. We consider reducing the number of communication rounds required for training the model by using additional computation, so as to achieve faster convergence. Therefore, we refer to FedADAM algorithm and propose a new adaptive method according to the idea of dynamic adjustment of static parameters such as learning rate and gradient.

The pseudocode of the Algorithm is presented in **Algorithm 1** and abstractly show in **Figure 2**:

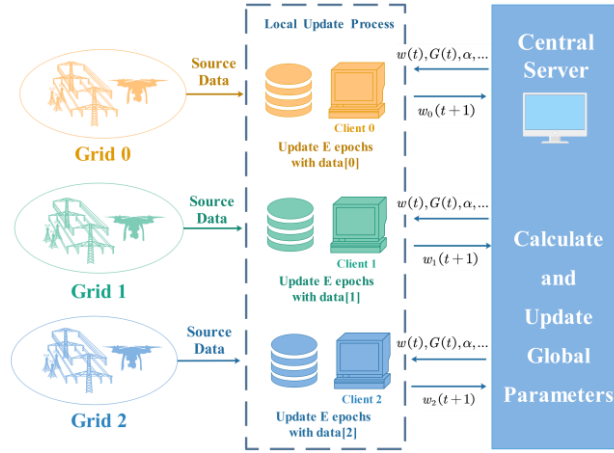


Figure 2 Adaptive Algorithm Process

Algorithm 1 Adaptive Algorithm

Preset: Hyper parameters $\tau, \alpha, \beta_1, \beta_2,$

initialize $w_0, v_{-1}, m_{-1} \geq \tau^2$

2: **for** $t = 0, 1, \dots, T - 1$ **do**

$K \leftarrow C \cdot |S|$

4: **for** each client $k = 0, 1, \dots, K - 1$ in parallel **do**

 Compute the gradient $\nabla F_k(w(t))$ of client k

6: $w_k(t) = w(t)$

$G_{k,0}(t) = G(t)$

8: **for** each epoch $e = 1, 2, \dots, E$ **do**

$G_{k,e}(t) = \alpha (\nabla F_{k,e}(w(t)) - G_{k,e-1}(t))$
 $+ (1 - \alpha) G_{k,e-1}(t)$

10: $w_{k,e}(t+1) = w_k(t) - \eta G_{k,e}(t)$

$e \leftarrow e + 1$

12: **end for**

$\Delta_k(t) = w_{k,E}(t+1) - w(t)$

14: **end for**

$w(t+1) = \sum_{k=0}^{K-1} \frac{D_k}{D} w_{k,E}(t+1)$

16: $G(t+1) = \frac{w(t) - w(t+1)}{\eta}$

$\Delta(t) = \sum_{k=0}^{K-1} \frac{D_k}{D} \Delta_k(t)$

18: $m_t = \beta_1 m_{t-1} + (1 - \beta_1) \Delta(t)$

$v_t = \beta_2 v_{t-1} + (1 - \beta_2) \Delta^2(t)$

20: $w(t+1) = w(t) + \eta \frac{m_t}{\sqrt{v_t + \tau}}$

end for

Algorithm 1 shows our adaptive approach, C is the fraction of clients participating in each round, S is the set of all clients, $\tau, \alpha, \beta_1, \beta_2$ are hyperparameters of the model which could be customize before training, η denotes the learning rate. Therefore, $K = C \cdot |S|$ clients are sampled at the beginning of each global iteration. Assume the local parameter that the k -th involved client received as $w_k(t)$, in each iteration the global parameter $w(t)$ is transferred thus $w_k(t)$ is assigned as:

$$w_k(t) = w(t)$$

Since the historical gradient information hasn't been recorded in the first global iteration, the gradient of the loss function $\nabla F_k(w(t))$ would be applied as the replacement of estimated local gradient function in the first epoch $G_{k,0}(t)$:

$$G_{k,0}(t) = \nabla F_k(w(t)), \text{ where } t = 0$$

As the gradient of the loss function $\nabla F_k(w(t))$ of participated clients differ from each other, which leads to a situation that FedAVG may have more divergent results and poor convergence when it performs multiple local updates for non-IDD dataset. To mitigate the unstable convergence performance of the model for non-IDD data in practical problems, the model could be adaptively adjusted by introducing the estimated gradient function, which enables the model to achieve better results for non-IDD data after multiple local updates in the correct update direction. After the first local update, the central server calculates estimated global gradient function $G(t+1)$ with the global parameters of current and the previous global iterations as:

$$G(t+1) = \frac{w(t) - w(t+1)}{\eta}$$

where $t = 0, 1, \dots, T-1$, which will be delivered to the participated clients in the next global iteration for the update of the estimated gradient function of client k at the remaining global iterations:

$$G_{k,0}(t) = G(t), \text{ where } t = 1, 2, \dots, T-1$$

As shown in Algorithm 1, the core idea of our adaptive method is to perform the update of $G_{k,e}(t)$ in each epoch e of client k as follows:

$$G_{k,e}(t) = \alpha(\nabla F_{k,e}(w(t)) - G_{k,e-1}(t)) + (1-\alpha)G_{k,e-1}(t)$$

where $\nabla F_{k,e}(w(t))$ is the gradient of the loss function with the local dataset D_k at each epoch, α is a hyper-parameter that we preset for adjusting $G_{k,e}(t)$. Then the update of local parameters at each epoch can be formulated as:

$$w_{k,e}(t+1) = w_k(t) - \eta G_{k,e}(t)$$

In addition, in order to accelerate the convergence speed and improve the performance of the model for non-IDD data, E -th epoch value of local parameters would be used to calculate $\Delta_k(t)$ according to the idea of Adam [7]:

$$\Delta_k(t) = w_{k,E}(t+1) - w_k(t)$$

During the aggregation phase, the central server receives the parameters $\Delta_k(t)$ and $w_{k,E}(t+1)$ after E local updates from the k involved clients, and aggregates their values to calculate the weighted average of the local

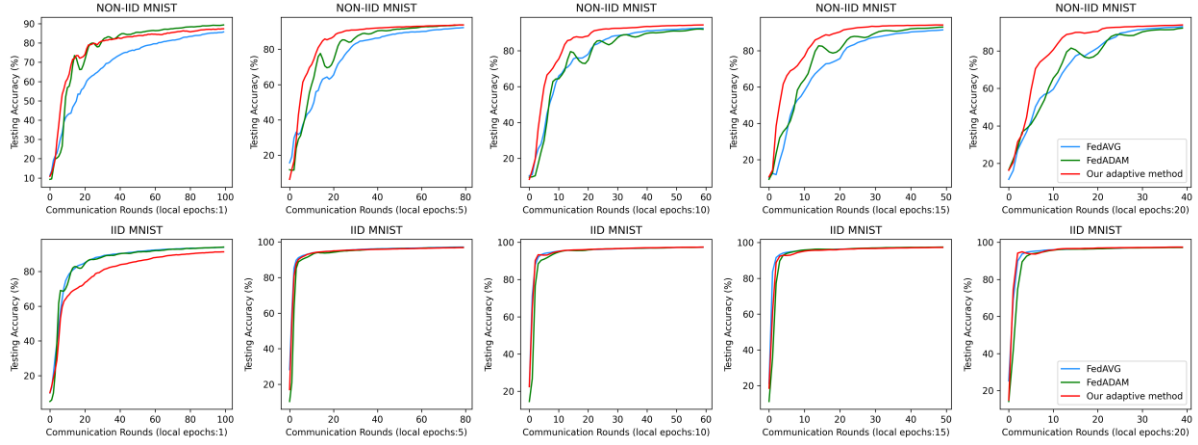


Figure 3 Testing accuracy of FedAVG, FedADAM, Our Adaptive Method over MNIST dataset, with non-IID, IID data

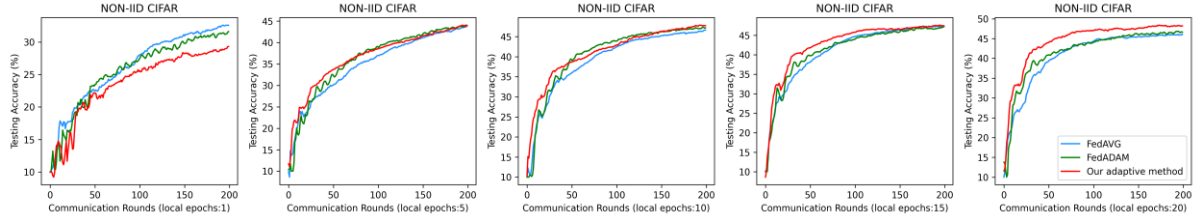


Figure 4 Testing accuracy of FedAVG, FedADAM, Our Adaptive Method over CIFAR-10 dataset, with non-IID data

parameters of the involved clients, which can be formulated as:

$$w(t+1) = \sum_{k=1}^K \frac{D_k}{D} w_{k,E}(t+1)$$

$$\Delta(t) = \sum_{k=1}^K \frac{D_k}{D} \Delta_k(t)$$

their values to calculate the weighted average of the local parameters of the involved clients, which can be formulated as:

At the end of each global iteration, the first moment estimation m_t and second moment estimation v_t are calculated with Adam's method, and finally the global parameters is updated by:

$$w(t+1) = w(t) + \eta \frac{m_t}{\sqrt{v_t + \tau}}$$

Such a design has the following advantages:

a) Source Data in most of the practical problems (e.g. Electric power inspection with UAV System in our background) is often distributed inconsistently, which makes the model trained at the case of non-IID data. By utilizing the global gradient information to adaptively update local parameter, faster and more robust convergence can be achieved.

b) According to the idea of FedADAM, we calculated the moment estimation using gradient information recorded in the history to adaptively regulate learning rate of training. It should receive much better effect in accelerating network training and inhibiting data oscillations.

c) The adaptive adjustment of parameters such as gradient and learning rate does not involve direct transmission of source data, which ensures data privacy of participants in federated learning.

d) Our adaptive adjustment focuses on the optimization in case of non-IID data, in which participated clients can obtain more stable parameters through more local training. Through this way,

the whole model is more inclined to train on the client side, which will greatly reduce the communication rounds and lower the cost of communication between the server and client.

3. Experiment and Comparison

In the following, we will compare Our Adaptive Method, FedAVG and FedADAM to show the superiority of Our Method under local multi-epoch training.

There is no publicly available data about the electric power inspection with UAV on the Internet at present. Without loss of generality, we selected MNIST [10] datasets and CIFAR-10 [11] datasets which are widely used in machine learning area for simulation respectively. We divide the dataset into IDD data and non-IDD data, construct the CNN model, and select different combinations of local epochs and communication round for training.

In the experiment, we set the learning rate $\eta=0.01$, and set $\alpha=0.03$ for local update of participants. At the same time, both FedADAM and Our Adaptive Method use hyper-parameters $\tau=0.001, \beta_1=0.9, \beta_2=0.99$ for training. As shown in **Figure 3** and **Figure 4**, Our Adaptive Method can achieve the expected result.

By analyzing the results of the algorithm, the following conclusions can be obtained:

a) For IDD data and implementing only one local training, the performance of Our Adaptive Method is not necessarily better than generic FedAVG and FedADAM, or even more mediocre. The reason is the main purpose of our adaptive improvement is to optimize the convergence speed and robustness of the model under the condition of non-IDD data. To save communication cost, more local training epochs is required to run on client. Under this circumstance, this method can truly attribute much faster and robust convergence with non-IDD data and significantly reduce the communication cost between server and client at the meantime, which is more meaningful in solving practical problems.

b) As the local epochs gradually increase, it can be observed that both FedAVG and FedADAM slow down the convergence speed to different degrees, while Our Adaptive Method keeps stable convergence performance. This is because there are often large differences between local data among participants for non-IID dataset. If the participated clients have gone through local updates for many times, the differences between participating clients will become larger and larger, thus the convergence efficiency will slow down a lot during server aggregation. However, Our Adaptive Method ensures higher accuracy and faster convergence speed through adaptive adjustment of parameters in this case.

4. Conclusion

In this article, we propose an adaptive FL method by using the momentum and adaptive gradient to optimize the convergence performance of the model. To achieve fast convergence, we introduce the new local gradient by considering the difference between the local gradient and historical global gradient. Furthermore, by tracking the first and second moment estimation of the gradients for model parameter, our algorithm adjusts the learning rate adaptively. At last, we perform simulation experiments using MNIST dataset and CIFAR-10 dataset to verify that our model can achieve faster convergence speed and higher accuracy than the widely used FL algorithm for non-IDD data. It is of great significance to accelerate the convergence of the model and reduce the communication cost in the practical problems of imbalanced data distribution.

5. Acknowledge

This paper is funded by *the Innovation and Entrepreneurship Training Program for College Students of Sun Yat-sen University*. (Project number: 202211500)

6. Reference

- [1] S. C. Sun and L. Y. Zhang, et al, " Application of intelligent identification technology in UAV power inspection," *The Journal of New Industrialization*, 2020.
- [2] B. McMahan and E. Moore, et al., "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*, pp. 1273-1282, 2017.
- [3] X. Li and K. Huang, et al., " On the Convergence of FedAvg on Non-IID Data," in *International Conference on Learning Representations*, 2020.
- [4] C. Xu and S. Liu, et al., "Learning rate optimization for federated learning exploiting over-the-air computation," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 12, pp. 3742-3756, 2021.
- [5] M. Duan and D. Liu, et al., "Self-balancing federated learning with global imbalanced data in mobile systems," *IEEE Transactions on Parallel and Distributed Systems*, vol. 32, no. 1, pp. 59-71, 2021.
- [6] S. Reddi and Z. Charles, et al., "Adaptive federated optimization," in *International Conference on Learning Representations*, 2021.
- [7] Kingma, D. P. and Ba, J., "Adam: A method for stochastic optimization," in *International Conference on Learning Representations*, 2015.
- [8] Hsu and T. M. H., et al. "Measuring the effects of non-identical data distribution for federated visual classification." *arXiv preprint arXiv:1909.06335*, 2019.
- [9] Mills, J and Hu, J., et al. "User-oriented multi-task federated deep learning for mobile edge computing." *arXiv preprint arXiv: 2007.09236*, 2020.
- [10] Y. LeCun and C. Cortes, "MNIST handwritten digit database," 2010. [Online]. Available: [MNIST handwritten digit database, Yann LeCun, Corinna Cortes and Chris Burges](#)
- [11] A. Krizhevsky, V. Nair, and G. Hinton, "Cifar-10 (Canadian institute for advanced research)," 2010. [Online]. Available: [CIFAR-10 and CIFAR-100 datasets \(toronto.edu\)](#)