An Adaptive Federated Control Strategy for Participant Selection in Multi-Client Collaboration

1st Yizhu Chen
State Key Laboratory of Management and Control for Complex Systems
Institute of Automation, Chinese Academy of Sciences
Beijing, China
School of Artificial Intelligence
University of Chinese Academy of Sciences
Beijing, China
chenyizhu2019@ia.ac.cn

2nd Xiaoming Du
School of Economics and Management
University of Chinese Academy of Sciences
Beijing, China
duxm@cashq.ac.cn

3rd Xiao Wang
State Key Laboratory of Management and Control for Complex Systems
Institute of Automation, Chinese Academy of Sciences
Beijing, China
Qingdao Academy of Intelligent Industries
Qingdao, China
x.wang@ia.ac.cn

4th Jun Jason Zhang
School of Electrical Engineering and Automation
Wuhan University
Wuhan, China
jun.zhang.ee@whu.edu.cn

5th Fei-Yue Wang*
State Key Laboratory of Management and Control for Complex Systems
Institute of Automation, Chinese Academy of Sciences
Beijing, China
feiyue.wang@ia.ac.cn

***Abstract*—The federated ecology provides a new paradigm for breaking the isolated data island problem and fully activating the potential of big data and artificial intelligence, especially in multi-client collaboration tasks. Participant selection strategy in multi-client collaboration is the main limitation for increasing the convergence speed and lowing the communication costs. However, in the face of the unbalanced and non-IID data distributions, the performance of federated optimization algorithms will also decrease. To solve the above problems, we propose an adaptive federated control strategy for participant selection in multi-client collaboration based on the *Mann-Kendall test*, named *FedMK*. By adaptively selecting weak participants for training rather than random selection, *FedMK* can speed up model convergence and reduce communication costs. Experiment results show that our method outperforms the baseline method in non-IID scenarios, and the number of communication rounds on CIFAR-10 and synthetic datasets reduced by more than 15% and 10%, respectively.**

***Index Terms*—federated control, non-IID, participant selection, Mann-Kendall test**

I. Introduction

With the rapid development of big data and artificial intelligence, modern society has put forward new requirements for data privacy protection and data security. However, in many application areas, data usually exists in a distributed form and is difficult to circulate effectively and ensure data privacy [1], [2]. In response to the above problems, *federated learning* [3] provides a solution for multi-client distributed training based on data privacy preservation. Clients learn the global model together in a collaborative way, and the data is only kept locally on the client and not shared. In the federated ecology proposed by Wang et al. [4], federated control is responsible for providing a regulatory mechanism for data circulation. One of the important issues in federated control is how to select preferred participants dynamically.

At present, most methods randomly select clients in each training round. In order to enable the federated learning model to converge faster and thus reduce the number of communication rounds, we propose an adaptive federated control strategy for participant selection in multi-client collaboration based on the *Mann-Kendall (MK) test* method, named *FedMK*. By detecting the prediction quality of each client in the recent rounds, *FedMK* can obtain a preferred list by selecting weak participants preferentially in each global round so that they have more local training opportunities. We further conducted experiments in two different datasets under non-independent and identically distributed (non-IID) settings. The results proved that our method could get a resulting model with higher performance in a shorter time.

We make the following contributions:

1) We are the first to use the trend detection method *Mann-*

The work is funded by the National Key R&D Program of China under Grant 2018AAA0101502 and the Science and Technology Project of SGCC (State Grid Corporation of China) : fundamental theory of human-in-the-loop hybrid-augmented intelligence for power grid dispatch and control.
*Corresponding author: Fei-Yue Wang.

2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence (DTPI) | 978-1-6654-3337-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/DTPI52967.2021.9540128

Kendall test for time series to assist in clients' predictive quality detection.

2) We propose an adaptive federated control strategy *FedMK* for the participant selection step. Our method can reduce the number of communication rounds with only a slight increase in computing costs.
3) We conducted a series of experiments on non-IID settings using both real and synthetic datasets and proved that our method is of generality.

II. Related Work

A. Federated Learning

The purpose of federated learning is to solve the risks and responsibilities of the centralized storage of data. Compared with traditional distributed learning, a fundamental change in federated learning is to transmit model parameters instead of data in a secure and encrypted way. Federated learning faces some new challenges in the statistical diversity of data. The data of each participant in federated learning may be non-IID distribution, and they may have datasets of different sizes. In addition, communication cost is also one of the challenges faced by federated learning. The existing research mainly focuses on compression models [5], increasing client computing costs [6], and optimizing participant selection methods to reduce communication costs and improve communication efficiency. In terms of improving participant selection methods, Nishio et al. [7] proposed a new federated learning protocol FedCS, which sets the deadline for each client to download, update and upload the local model. Yoshida et al. [8] proposed a hybrid learning mechanism Hybrid-FL to solve client- and data-selection problems. In this paper, we improve the participant selection step by combining historical detection trends of clients.

B. Mann-Kendall Test

Mann-Kendall (MK) test is a trend detection method of time series data. Its characteristic is that samples do not need to follow a specific distribution and are affected by a small number of outliers. For time series $X = (X_1, X_2, ..., X_n)$, *MK test* can determine whether to accept null hypothesis H_0 or alternative hypothesis H_1,

$$H_0 : X \text{ has no monotonic trend.}$$
$$H_1 : X \text{ has monotonic trend.}$$

If the null hypothesis H_0 is rejected, it means that at the confidence level α, there is a clear upward or downward trend in this time series. According to the test statistics Z, it can be determined whether the series is increasing or decreasing.

III. Methods

We propose an adaptive federated control strategy for participant selection in multi-client collaboration, *FedMK*, based on the *MK test*. In this section, we will first introduce the preliminaries of our proposal and then give the overall process of *FedMK*. Finally, we will introduce the participant selection strategy in detail.

Algorithm 1 *FedMK*: An adaptive federated control strategy for participant selection. E is the local epochs, B is the local batch size and l is the length of the historical list.

Server:

```
initialize E, l, w̄^0, A^0
for each round t = 1, 2, ... do
    Cs^t ← Participant list from Algorithm 2
    for each client k ∈ Cs^t do
        w_k^t, A_k^t ← Client(w̄^{t-1})
        A^t ← Update historical list A^{t-1}
    end for
    w̄^t ← Σ_{k∈S_t} (n_k/n) w_k^t
end for
```

Client: Run on client k

```
for r = 0 to E − 1 do
    for each mini-batch D_i with size B do
        Update w_k^t by minimizing L(w_k^t | w̄^{t-1}, D_i)
    end for
end for
return w_k^t and test accuracy A_k^t
```

A. Preliminaries

Federated learning algorithms involve large-scale clients' collaborative training scenarios. Specifically, the goal is to optimize the following objective function

$$f(w) = \sum_{k=1}^{K} \frac{n_k}{n} F_k(w) \text{ where } F_k(w) = \frac{1}{n_k} \sum_{i\in\mathcal{P}_k} f_i(w) \quad (1)$$

where K is the total number of clients, $f_i(\cdot)$ denotes the loss function. The dataset in the k-th client is $\mathcal{P}_k$ and $n_k = |\mathcal{P}_k|$ represents the number of samples, $n = \sum_k n_k$ is the total number of samples in all clients.

FedAvg is one of the earliest methods to solve (1). The method runs simply by selecting a certain proportion of devices and allowing each selected device to run local stochastic gradient descent (SGD) of E epochs. Then the server calculates the average of the local models to obtain the resulting models. The process of local training and global averaging is carried out iteratively throughout the federated learning task. In this work, *FedAvg* is utilized as the baseline method in the non-IID scenarios.

B. FedMK: Participants Adaptively Selection Algorithm

We propose a variant of *FedAvg* named *FedMK* that was based on *MK test* to focus on the learning process of weak clients adaptively. Since all the problems involved in our research are multi-classification problems, we use the cross-entropy loss function and calculate it as

$$F_k(w) = -\frac{1}{n_k} \sum_{i=1}^{n_k} \sum_{j=1}^{C} y_{i,j} \log p_{i,j} \quad (2)$$

where $p_{i,j}$ is the probability that the i-th sample is predicted to be the j-th label value and $y_{i,j}$ is the true label. The objective of each client is to minimize (2).

In order to reduce total training rounds, our method will pay more attention to the clients with relatively weak performance in recent historical rounds. By giving priority to these clients in the next round of training, they could have more opportunities for local training to improve the convergence speed of the resulting model. Based on the *MK test*, we now propose our adaptive participant selection algorithm *FedMK*. Here, we use the *MK test* to evaluate the predictive quality of each client. We maintain a historical list of length l for each client in the server, which stores the test accuracy of in recent l rounds; then, we use the *MK test* to determine which clients have a decreasing historical accuracy trend. If the client shows a decreasing trend, it will be added to the candidate list Ca^t and be selected firstly in the next round participant selection step. After evaluating all clients, we will get the participant list for the t-th round.

The overall training process of *FedMK* is similar to *FedAvg*. At the t-th iteration, the server selects the participant list according to our adaptive participant selection step based on the *MK test*; then, the server sends the global model to all selected participants. After receiving the global model, the client executes local training for E epochs on the local data and uploads the updated model. In order to select a preferred participant list, we will upload the test accuracy of selected participants to the server additional. The server will update the historical information of each client as the basis for the next participant selection step. Finally, the server performs a model aggregation operation to generate the global sharing model. See Algorithm 1 for the complete *FedMK* process.

Our adaptive federated control strategy *FedMK* makes the training process more efficient by paying more attention to the clients whose recent predictive quality is decreasing. In addition, it only occupies minimal server resources and time, and the computational costs are also kept minimal.

C. Participant Selection Step with MK Test

Now we introduce how to use *MK test* to evaluate the predictive quality of clients. Through the historical list maintained by the server, we can get the recent detection trend of each client using the *MK test*. After that, clients showing a decreasing trend will be added to the candidate list Ca^t.

Since the length of the Ca^t is uncertain, we distinguish the following two cases to design different selection strategies (we use $N = |S_t|$ to represent the number of clients that need to be selected in each round, and use Cs^t to represent the participant list selected in round t): i) if $|Ca^t| \leqslant N$, then add all clients from Ca^t to Cs^t and select $N - |Ca^t|$ participants from the remaining clients randomly; ii) if $|Ca^t| > N$, then select N participants from Ca^t as Cs^t randomly. The list Cs^t will participate in the $t+1$ training round. Our adaptive participant selection step is described as Algorithm 2.

Algorithm 2 Participant selection step based on *MK test*.

Input: Clients' historical list A^{t-1}, number of clients K, confidence α, number of selected participants N and current global training round t

Output: List of selected participants Cs^t

Initialize $Cs^t = [\ ]$
for $k = 0$ to $K-1$ **do**
  Calculate statistic S for list $A^{t-1}[k]$:
$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sgn(x_j - x_i)$$
  Calculate $Var(S)$:
$$Var(S) = \frac{1}{18}[n(n-1)(2n+5) - \sum_{p=1}^{g} t_p(t_p-1)(2t_p+5)]$$
  Calculate statistic Z:
$$Z = \begin{cases} (S-1)/\sqrt{Var(S)} & S > 0 \\ 0 & S = 0 \\ (S+1)/\sqrt{Var(S)} & S < 0 \end{cases}$$
  if $|Z| \geq Z_{1-\alpha/2}$ **and** $Z > 0$ **then**
    Add client k to alternative list Ca^t
  else continue
end for
if $len(Ca^t) > N$ **then**
  $Cs^t \leftarrow$ Choose N participants from Ca^t randomly
else
  $Ct^t \leftarrow$ Choose $N - len(Ca^t)$ participants from the remaining clients randomly
  $Cs^t \leftarrow [Ca^t, Ct^t]$
return Cs^t

IV. Experiments and Results

We consider a classification problem using both real dataset CIFAR-10 and synthetic dataset Synthetic. Our method is based on *FedAvg* to improve the participant selection step, so we choose *FedAvg* as the baseline method. We will show that our adaptive participant selection method for multi-client collaboration in federated control performs better in model convergence speed, effectively reducing the communication costs during the training process.

A. FedMK on CIFAR-10

The real dataset CIFAR-10 [9] is composed of 10 classes of 32×32 images with three RGB channels. Due to the limitation of the size, we distribute CIFAR-10 to $K = 20$ clients. In order to simulate the non-IID settings, we refer to the settings in [10] to simulate 20 clients participating in the federated learning task. Each client only contains 3 of the 10 classes.

We use the same model as *FedFusion* [11], which has two 5×5 convolution layers (both has 64 channels and each followed by a ReLU activation and 2×2 maximum pooling), two fully connected layers (the first has 384 units and the second has 193 units, both followed by a ReLU activation and random dropout) and a softmax output layer. For a convenient and fair comparison, we selected a fixed set of hyperparameters, that is, 5 participants selected in each global round, with the learning rate of $\eta = 3 \times 10^{-2}$, the local batch size of $B = 64$, and the local epochs of $E = 20$.

The experimental results on CIFAR-10 are shown in Fig.1. Specifically, *FedMK* reached the test accuracy of 60% by

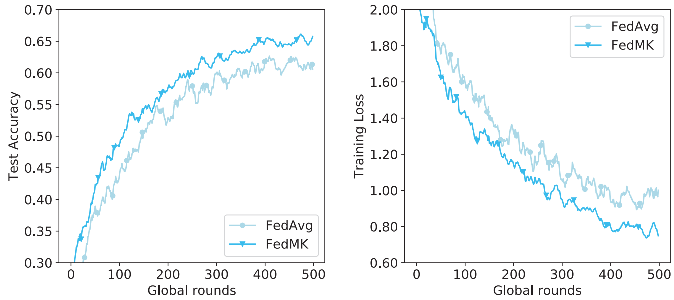

Fig. 1. Results on CIFAR-10. The left accounts for test accuracy vs. global rounds while the right for training loss vs. global rounds.

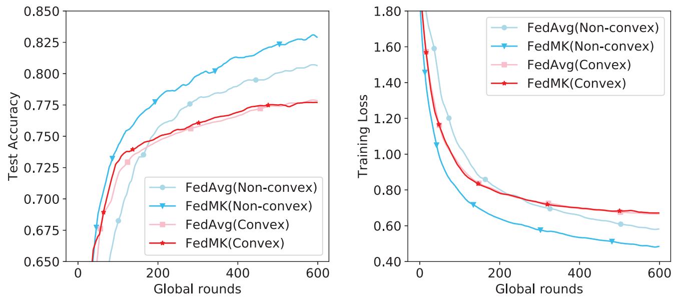

Fig. 2. Results on Synthetic. The left accounts for test accuracy vs. global rounds while the right for training loss vs. global rounds.

193 rounds, achieving a reduction of 15.7% than the baseline method. The test accuracy of our method reached 66.12% finally, which was 4.36% higher than the baseline method.

B. *FedMK on Synthetic*

In order to more accurately model the statistical heterogeneity of different clients, we also evaluate our model in a synthetic dataset. The settings for generating the synthetic dataset are similar to Shamir et al. [12]. Refer to [13], we control the data generation and distribute process through parameters $\bar{\alpha}$ and $\bar{\beta}$ to enhance the degree of heterogeneity among clients. The dataset finally contains 10 classes of 60-dimensional real-valued data. We set $\bar{\alpha} = 0.5$ and $\bar{\beta} = 0.5$. The data size of each client follows the power law and is distributed in [250, 25810]. We generate $K = 100$ clients and select $N = 10$ clients in each global round.

As for the synthetic dataset, we use convex and non-convex models for training. We use the multinomial logistic regression model (MLR) with L2 norm in the convex settings. In the non-convex settings, we use a two-layer deep neural network (DNN) with a hidden layer size of 20 (the first fully connected layer is followed by a ReLU activation, and the network is connected to a softmax layer finally). Subsequent experiments will use the same batch size of $B = 20$, local iterations of $E = 20$ to compare different algorithms. And we set learning rate $\eta = 5 \times 10^{-3}$ for convex settings and $\eta = 3 \times 10^{-2}$ for non-convex settings.

The results on Synthetic are shown in Fig.2. For the convex MLR model, *FedMK* reached the test accuracy of 77% by 407 rounds, achieving a reduction of 11.1% than the baseline method. The final test accuracy is slightly higher than the baseline. Moreover, for the non-convex DNN model, *FedMK* reached the test accuracy of 80% through 312 rounds. The communication rounds to reach this performance achieved a reduction of 38.0% than the baseline method. The test accuracy of our method reached 83.19% finally, which was 2.33% higher than the baseline.

V. Conclusion

In this work, we propose an adaptive federated control strategy for participant selection in multi-client collaboration based on the *Mann-Kendall test*, named *FedMK*. Our results show that introducing *FedMK* method can obtain a global model with higher precision and lower communication overhead, and can perform better than baseline in non-IID settings.

Although our method has achieved better results, we have not explored all the potential of *FedMK* algorithm, such as combining this adaptive selection strategy with other federated optimization methods. We hope to explore a broader range of federated control strategy in future work.

References

[1] T. Shen, J. Wang, C. Gou, and F.-Y. Wang, "Hierarchical fused model with deep learning and type-2 fuzzy learning for breast cancer diagnosis," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 12, pp. 3204–3218, 2020.

[2] Q. Wei, H. Li, and F.-Y. Wang, "Parallel control for continuous-time linear systems: A case study," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 4, pp. 919–928, 2020.

[3] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial Intelligence and Statistics*, pp. 1273–1282, PMLR, 2017.

[4] F.-Y. Wang, Y. Wang, Y. Chen, Y. Tian, H. Qi, X. Wang, W. Zhang, J. Zhang, and Y. Yuan, "Federated ecology: from federated data to federated intelligence," *Chinese Journal of Intelligent Science and Technology*, vol. 2, no. 4, p. 305, 2020.

[5] S. Caldas, J. Konečny, H. B. McMahan, and A. Talwalkar, "Expanding the reach of federated learning by reducing client resource requirements," *arXiv preprint arXiv:1812.07210*, 2018.

[6] J. Konečnỳ, "Stochastic, distributed and federated optimization for machine learning," *arXiv preprint arXiv:1707.01155*, 2017.

[7] T. Nishio and R. Yonetani, "Client selection for federated learning with heterogeneous resources in mobile edge," in *ICC 2019-2019 IEEE International Conference on Communications (ICC)*, pp. 1–7, IEEE, 2019.

[8] N. Yoshida, T. Nishio, M. Morikura, K. Yamamoto, and R. Yonetani, "Hybrid-fl for wireless networks: Cooperative learning mechanism using non-iid data," in *ICC 2020-2020 IEEE International Conference on Communications (ICC)*, pp. 1–7, IEEE, 2020.

[9] A. Krizhevsky, G. Hinton, *et al.*, "Learning multiple layers of features from tiny images," 2009.

[10] C. T. Dinh, N. H. Tran, and T. D. Nguyen, "Personalized federated learning with moreau envelopes," *arXiv preprint arXiv:2006.08848*, 2020.

[11] X. Yao, T. Huang, C. Wu, R. Zhang, and L. Sun, "Towards faster and better federated learning: A feature fusion approach," in *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 175–179, IEEE, 2019.

[12] O. Shamir, N. Srebro, and T. Zhang, "Communication-efficient distributed optimization using an approximate newton-type method," in *International conference on machine learning*, pp. 1000–1008, PMLR, 2014.

[13] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *arXiv preprint arXiv:1812.06127*, 2018.