연합학습에서 클래스 균형을 고려한 디바이스 스케줄링 알고리즘

김상묵, 배상민, 엄성하, 윤세영

한국과학기술원 김재철 AI 대학원

{sangmook.kim, bamn0223, doubleb, yunseyoung}@kaist.ac.kr

Device Scheduling Algorithm Considering Class Balance in Federated Learning

SangMook Kim, Sangmin Bae, Seongha Eom, Se-Young Yun

Kim Jaechul Graduate School of AI, KAIST

요 약

Federated Learning is a practical distributed learning framework that can cooperatively learn a shared global model without direct access to data on edge devices. In federated learning, only a part of the entire device participates in learning due to communication restrictions. However, in a federated learning environment where each device's data is generated internally, the distribution of the data of each device is all heterogeneous. The Non-IID (Non-Independent and Identically Distributed) characteristics of these devices not only hinder the convergence of the model but also cause degradation of the model's performance. In this paper, we propose a scheduling algorithm that makes the overall data distribution of devices participating in learning balanced. Our proposed scheduling algorithm is not only very stable in learning on CIFAR10/CIFAR100 data but also shows high accuracy compared to Random Sample and Round Robin Scheduling.

1. Introduction

Federated Learning is a framework in which multiple local edge devices and a central server cooperate to train a global model without storing local clients' data in the central server[1]. The main goal of federated learning is to solve the following optimization problem:

$$\min_{w \in \mathbb{R}^d} \frac{1}{K} \sum_{i=1}^{K} f_i(w) \tag{1}$$

$$f_i(w) = \mathbb{E}_{(x,y)\in D_i}[l_i(w; x, y)] \tag{2}$$

where w is the weights of the global model, and K is the total number of local clients. l_i and D_i are the loss function and local training dataset of the i-th client, respectively.

One of the main challenges in federated learning is statistical heterogeneity. Since the local data of all devices reflect users' personal preferences and characteristics, the data distribution between devices is heterogeneous. The statistical heterogeneity causes weight divergence [2] and the performance degradation of the global model.

We propose a scheduling algorithm, Re-Balancing Scheduling(RBS), that selects devices by considering class balance and privacy-preserving in the device selection process. As shown in Figure 1, the central server selects devices so that the overall class distribution is balanced. Our scheduling algorithm considering this class balance reduces the weight divergence to mitigate model divergence and performance degradation.

2. Re-Balancing Scheduling Algorithm

2.1 Federated Averaging

Federated Averaging(FedAvg) [1], which is considered de facto federated learning algorithm, follows a four-step learning pipeline. In the first step, a device to participate in learning is selected. In the second step, the central server sends the current global model to all selected devices. In the third step, each device updates the model with its local data. In the last step, the device sends the updated model to the central server, and the central server aggregates the received models. Our scheduling algorithm also follows the FedAvg pipeline and selects devices by considering class balance in the device selection stage.

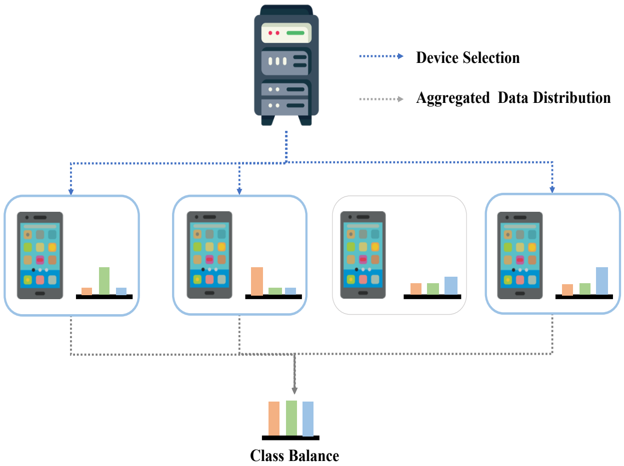


Figure 1. Device scheduling considering class balance. In the federated learning setting, our algorithm selects devices so that the class distribution of the group of devices is balanced.

Algorithm 1 Re-Balancing Scheduling Algorithm

α: Participation Ratio

K: The total number of devices

$R_{interval}$: Reset interval

1: **For** each communication round $t = 1, 2, \dots, T$ **do**
2: $\quad m \leftarrow \max(\alpha K, 1)$
3: $\quad$ **If** $t \,\%\, R_{interval} == 1$ **then**
4: $\qquad C \leftarrow \{1, 2, 3, \dots, K\}$
5: $\quad$ **end if**
6: $\quad$ Randomly select $i \in C$
7: $\quad C \leftarrow C - \{i\}$
8: $\quad S_t \leftarrow \{i\}$
10: $\quad$ **while** $|S_t| < m$ **do**
11: $\qquad k \leftarrow argmin_{k \in C} \sum_{i \in S_t} Sim(i, k)$
12: $\qquad S_t \leftarrow S_t \cup \{k\}$
13: $\qquad C \leftarrow C - \{k\}$
14: $\quad$ **end while**
15: # Perform Federated Learning Algorithm

2.2 Similarity of data distribution between devices

Re-balancing scheduling uses the data distribution similarity between devices. Recently, there have been attempts to calculate similarity without exchanging local data distribution information. Among them, we used the method suggested by [3]. [3] used the cosine similarity between predictions for the same random input (Gaussian Noise) $\tilde{x}$ in each local model. The data distribution similarity of devices i and j is calculated as follows:

$$Sim(i, j) = Cosin(w_i(\tilde{x}), w_j(\tilde{x})) \quad (3)$$

2.3 Re-Balancing Scheduling Algorithm

We propose a Re-Balancing Scheduling algorithm to mitigate model divergence and performance degradation. Algorithm 1 is our device scheduling algorithm that makes the entire class distribution of selected devices balanced. The intuition of our algorithm is to add a device with low similarity to the selected devices (lines 10-13). In this way, the device with many minor classes is selected to make the overall class distribution balanced. We allow a variety of devices to participate in learning by ensuring that selected devices are not re-selected for a specific communication interval (lines 3-5).

3. Experimental Results

3.1 Experimental Configuration

Datasets. We used CIFAR10 and CIFAR100 for performance verification. The sharding technique presented in [1] was used as a partitioning strategy for the Non-IIDness setting. After sorting the entire data by label class, it is divided into $s \times K$ shards. And allocate s shards to each device, where K is the total number of devices and s is the number of shards to be allocated. We set s to 2 and 20 for CIFAR10 and CIFAR 100, respectively.

Federated learning configuration. We set the total number of devices(K) to 100 and the participation ratio(α) to 0.1. We set the total communication rounds(T) to 300 and the local epochs to 5. We used SGD as the learning optimizer, and momentum and weight decay were set to 0.9 and 0.0001, respectively. We used a 4 Conv neural network model consisting of four convolution layers and one linear classifier.

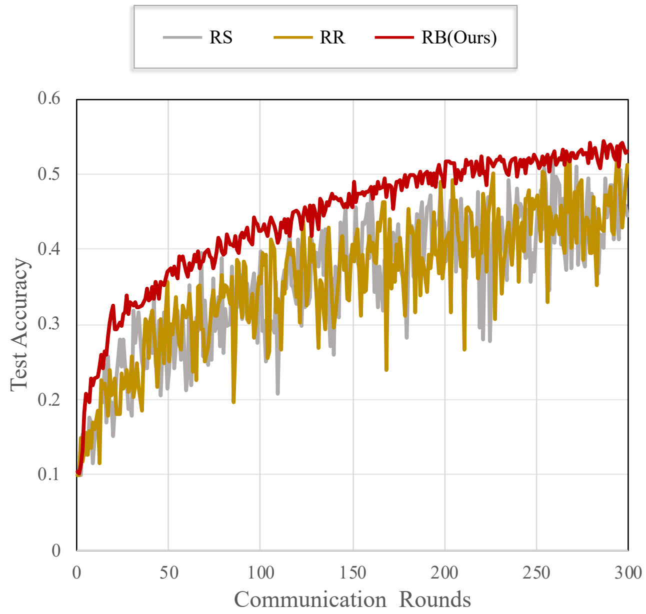
3.2 Performance Comparison

Two existing scheduling algorithms [4] were selected to compare with the performance of our proposed Re-Balancing (RB) scheduling algorithm. Random Sampling(RS) scheduling randomly selects local devices for each communication round. Round Robin(RR) scheduling arranges devices in groups and accesses groups consecutively.

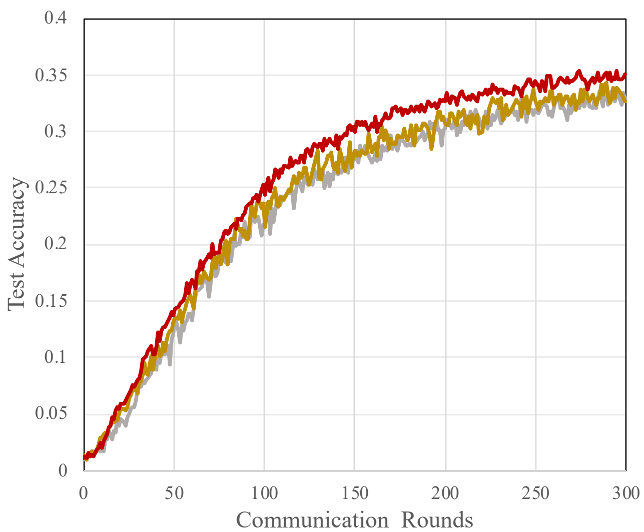
Scheduling	CIFAR 10	CIFAR 100
RS	52.56	33.61
RR	52.59	34.23
RB (ours)	54.48	35.34

Table 1. Test accuracy(%) comparison according to scheduling method with CIFAR 10 and CIFAR 100 dataset.

As can be seen from Table 1, our proposed scheduling method shows approximately 1.9% performance improvement at CIFAR 10 and 1.7% at CIFAR 100 compared to other scheduling methods. Figure 2 shows our algorithm is very stable and provides a reason for claiming that our algorithm mitigates model divergence and performance degradation by considering class balance.



(a) CIFAR 10



(b) CIFAR 100

Figure2. Test Accuracy vs. communication rounds for the CIFAR dataset.

4. Conclusion

Data heterogeneity between local devices causes model divergence and performance degradation in federated learning. Our proposed Re-Balancing Scheduling algorithm selects devices so that the entire class distribution is balanced during the device selection process. We show that our algorithm is stable and that performance gain can be obtained on the CIFAR10/CIFAR100 dataset. We leave the convergence analysis of this algorithm as future work.

참고 문헌

[1] Brendan McMahan, Eider Moore, Daniel Ramage, SethHampson, and Blaise Agueray Arcas. Communication-efficient learning of deep networks from decentralized data. In Artificial Intelligence and Statistics, pages 1273–1282.PMLR, 2017.

[2] Yue Zhao, Meng Li, Liangzhen Lai, Naveen Suda, Damon Civin, Vikas Chandra. Federated Learning with Non-IID Data. arXiv preprint arXiv:1806.00582.

[3] Wonyong Jeong, Jaehong Yoon, Eunho Yang, Sung Ju Hwang. Federated Semi-Supervised Learning with Inter-Client Consistency & Disjoint Learning. In International Conference on Learning Representations, 2021.

[4] Howard H. Yang; Zuozhu Liu; Tony Q. S. Quek; H. Vincent Poor. Scheduling Policies for Federated Learning in Wireless Networks. IEEE Transactions on Communications Volume: 68, 2020.

Acknowledgement

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2019-0-00075, Artificial Intelligence Graduate School Program(KAIST), 10%) and the Institute of Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (No. 2022-0-00871, Development of AI Autonomy and Knowledge Enhancement for AI Agent Collaboration, 90%)