◀ BIT ▶

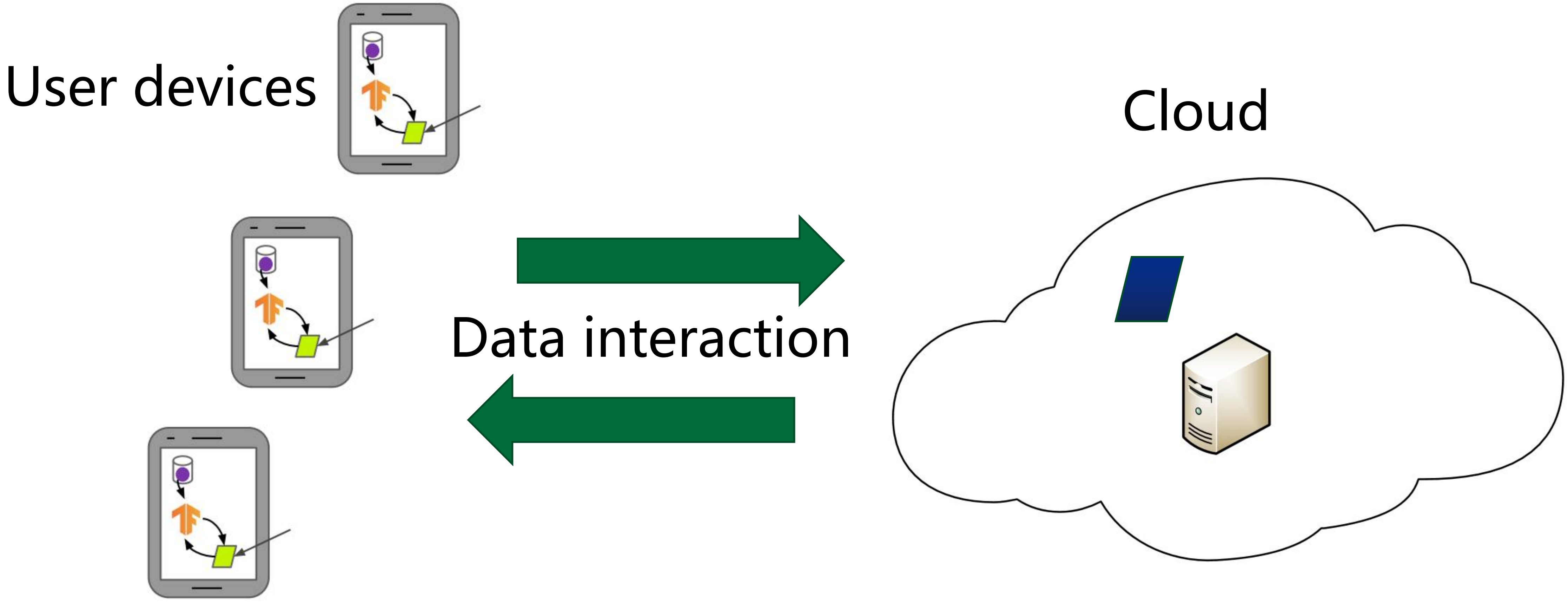
Maintaining Training Efficiency and Accuracy for Edge-assisted Online Federated Learning with ABS

Jiayu Wang, Zehua Guo, Sen Liu, Yuanqing Xia
Beijing Institute Of Technology, Fudan University

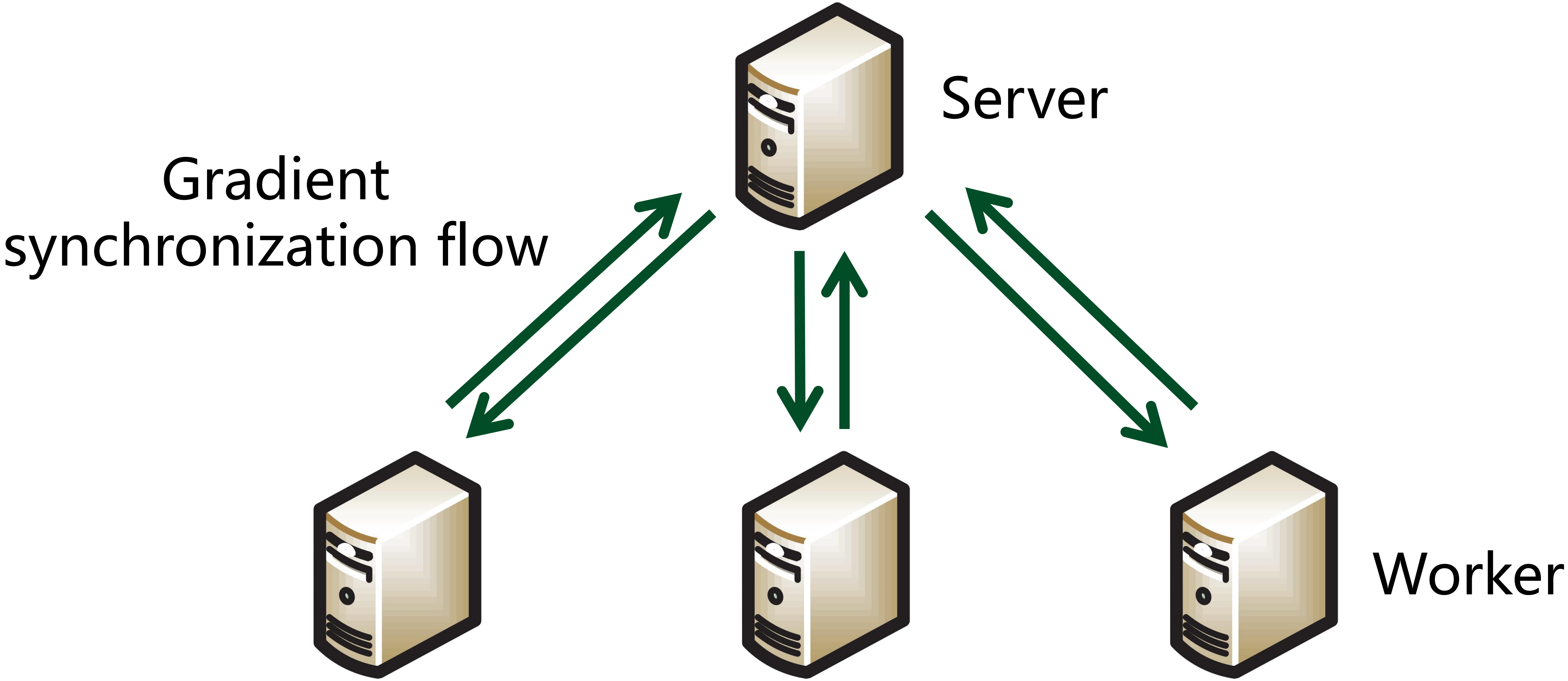
德以明理 学以精工

北京理工大學
BEIJING INSTITUTE OF TECHNOLOGY

Federated Learning



Parameter Server



Existing method

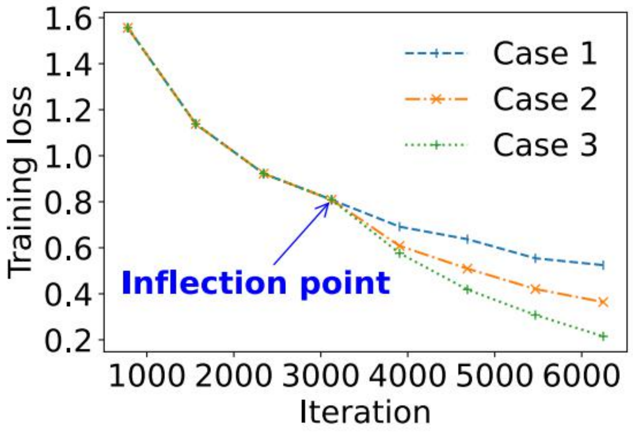
	Existing method	Existing problem
Training batch size	Training data batch size can fluctuate.	The decrease in batch size can have a negative effect on the training process.
Computing speed	Do not consider of the difference of computing speed.	Worker with more training data and low computing speed may drag the training process.
Utilization of the training data	Do not consider of the utilization of the training data.	The improper batch size can decrease the utilization of the training data.

Increase batch size

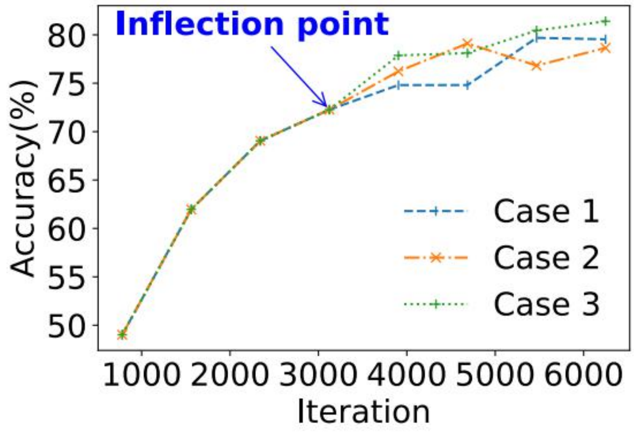
Training model: Resnet18
Dataset: CIFAR10

Iteration	Batch size
Case1	32 to 32
Case2	32 to 64
Case3	32 to 128

Changing of batch size



(a) Training loss in the case of increased batch size.



(b) Accuracy in the case of increased batch size.

- Increase the batch size can accelerate the training process.
- More improvement can further accelerate the training

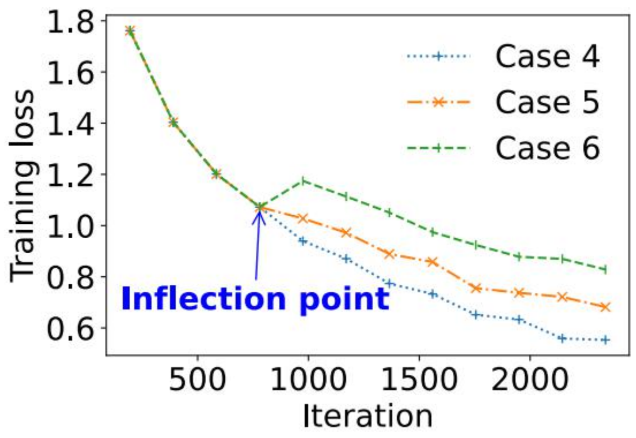
2 Observation

Decrease batch size

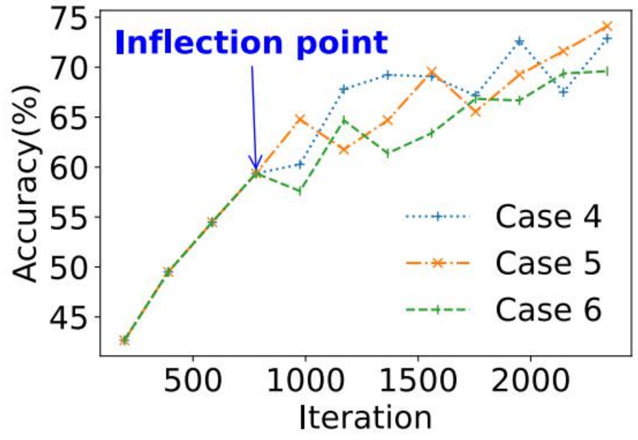
Training model: Resnet18
Dataset: CIFAR10

Iteration	Batch size
Case4	128 to 128
Case5	128 to 64
Case6	128 to 32

Changing of batch size



(c) Training loss in the case of decreased batch size.

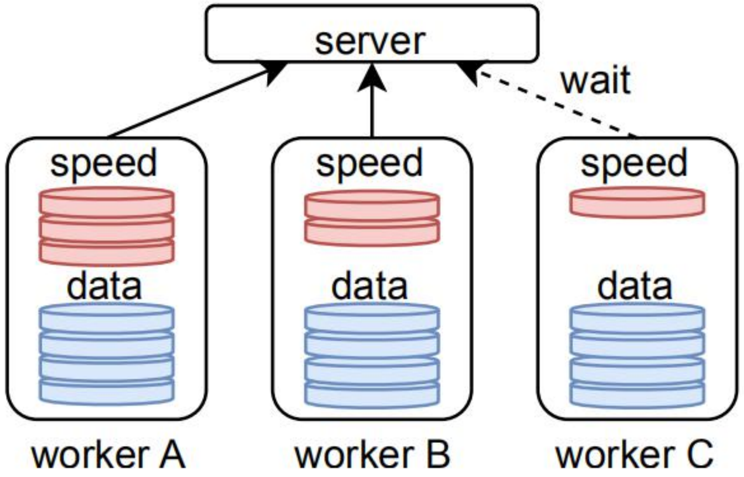


(d) Accuracy in the case of decreased batch size.

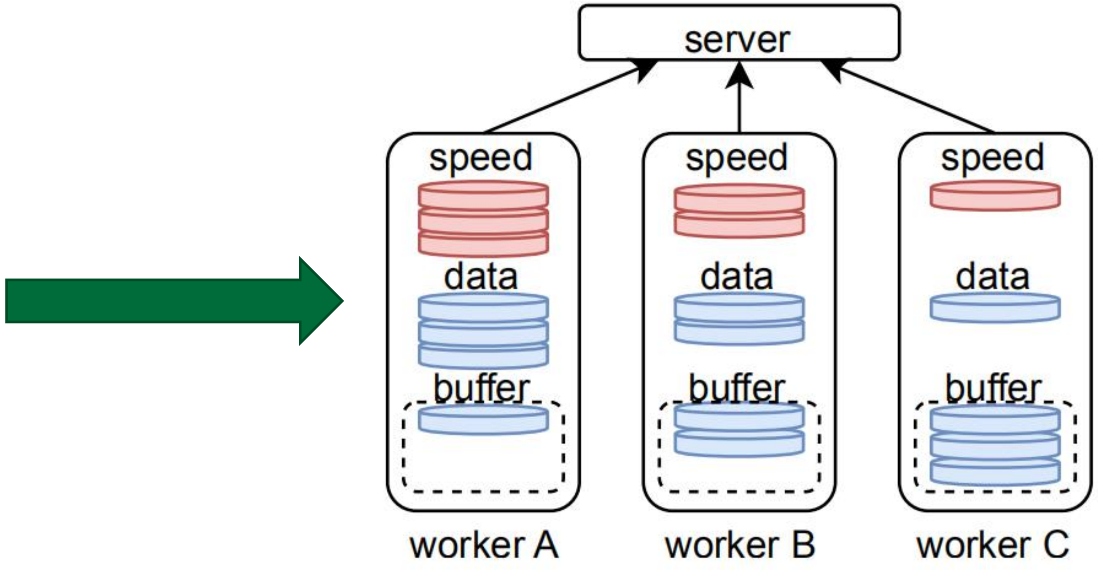
- A decrease in the batch size can slow down the training.
- Extreme small batch size will have a serious negative effect and lead to a long training process duration.

Our method

- Consider of the changeable data receiving speed, we adopt an adaptative batch size.
- Consider of the different computing speed, we set different batch size upper bound for different workers.
- To improve the utilization of the training data, we adpot lower bound for the training batch size.



Existing method



Our method

Warm-up phase

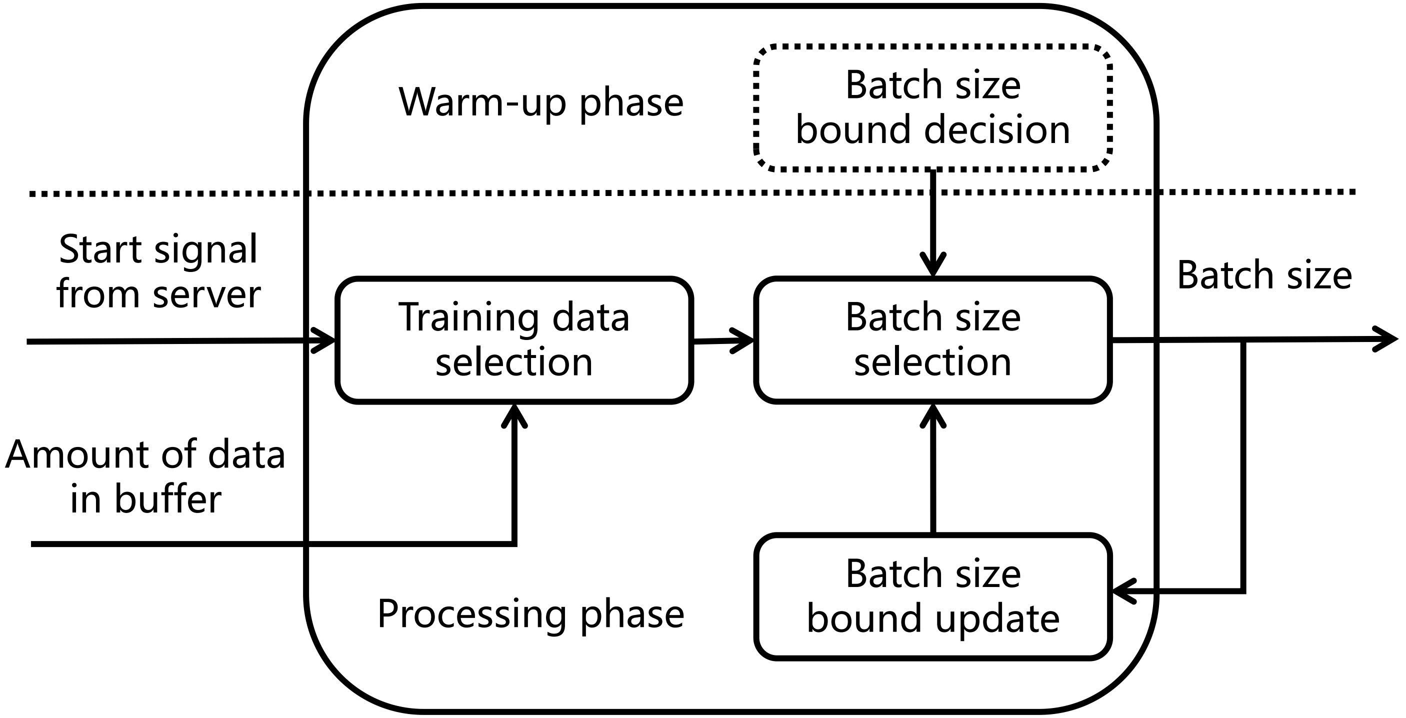
The setting of lower bound:

- We train the machine learning model with different batch size on the training data with one iteration, the batch size with the best training result will be set as the lower bound.

The setting of upper bound:

- We set an iteration duration at first. In each worker, the maximum batch size, which can be processed within this duration, will be set as the upper bound.

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY



ABS structure

■Processing phase

- **Training data selection:** Choose C% of the data.
- **Batch size selection:** Restrict the batch size within the bound.
- **Batch size bound update:** Compare the batch size with the lower bound and update the lower bound.

Experimental setup

Training data

- CIFAR10 dataset.

Training model

- Base on Resnet18 and adjust the last layer.

Simulation of the data stream

- We download the traffic dataset from Kaggle.

Other parameters

- We assume there is no network congestion.
- We choose 1% of the data in each iteration.
- We improve the lower bound when the training batch size is higher than the lower bound for 120 iterations.

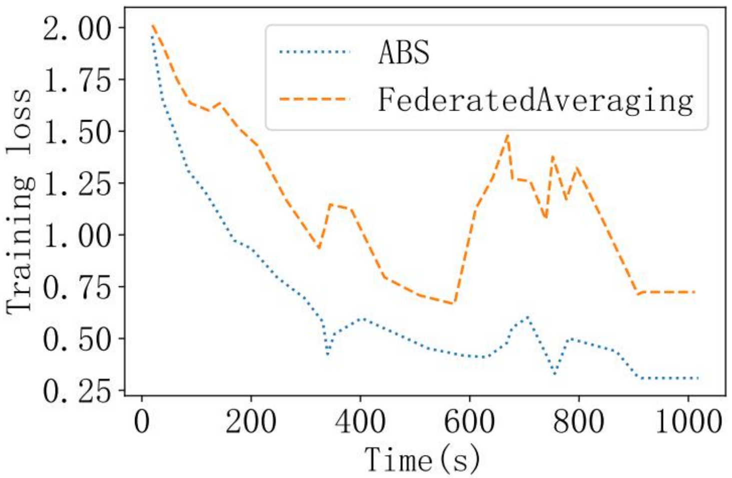
Comparison algorithm

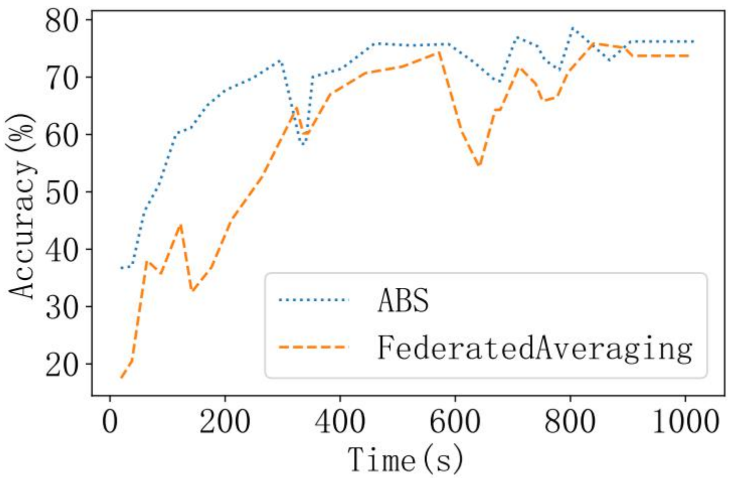
- FederatedAveraging: Each worker's training batch size is the size of all the data on it.

Experimental result

Training loss

Accuracy





- The training loss of ABS can convergence faster and more smooth.
- The testing accuracy of ABS can be higher.

北京理工大學

BEIJING INSTITUTE OF TECHNOLOGY

Thank you!

Questions?