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
Federated Learning

User devices ➔ Data interaction ➔ Cloud
Gradient synchronization flow
## Existing method

<table>
<thead>
<tr>
<th>Training batch size</th>
<th>Training data batch size can fluctuate.</th>
<th>The decrease in batch size can have a negative effect on the training process.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computing speed</td>
<td>Do not consider of the difference of computing speed.</td>
<td>Worker with more training data and low computing speed may drag the training process.</td>
</tr>
<tr>
<td>Utilization of the training data</td>
<td>Do not consider of the utilization of the training data.</td>
<td>The improper batch size can decrease the utilization of the training data.</td>
</tr>
</tbody>
</table>
Observation

**Increase batch size**

Training model: Resnet18
Dataset: CIFAR10

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Batch size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case1</td>
<td>32 to 32</td>
</tr>
<tr>
<td>Case2</td>
<td>32 to 64</td>
</tr>
<tr>
<td>Case3</td>
<td>32 to 128</td>
</tr>
</tbody>
</table>

Changing of batch size

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

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

**Decrease batch size**

Training model: Resnet18  
Dataset: CIFAR10

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Batch size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case4</td>
<td>128 to 128</td>
</tr>
<tr>
<td>Case5</td>
<td>128 to 64</td>
</tr>
<tr>
<td>Case6</td>
<td>128 to 32</td>
</tr>
</tbody>
</table>

- 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.

(c) Training loss in the case of decreased batch size.  
(d) Accuracy in the case of decreased batch size.
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 adopt lower bound for the training batch size.
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.
System design

- **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.

Diagram:
- Warm-up phase
  - Start signal from server
  - Amount of data in buffer
- 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.

ABS structure
## 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

- The training loss of ABS can converge faster and more smoothly.
- The testing accuracy of ABS can be higher.
Thank you!
Questions?