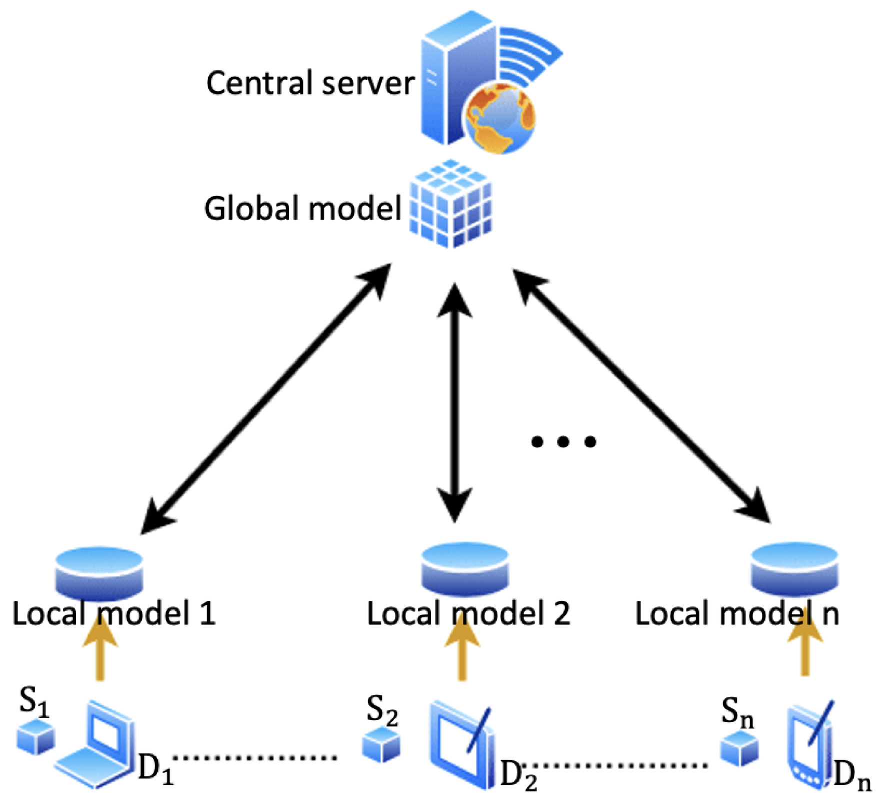
Semi-Synchronous Federated Learning with Adaptive Parameter Freezing

Zhou Rui

Supervisor: Dr C.H.Ngai

Introduction

Federated learning(FL) is developed for edge devices to train a global machine learning model collaboratively while keeping the data strictly to themselves.



- **Broadcast**
  Central server $\xrightarrow{w_t}$ local devices
- **Local update**
  $D_1 : w_t \rightarrow w_t^1$
  $D_2 : w_t \rightarrow w_t^2$
  ...
  $D_n : w_t \rightarrow w_t^n$
- **Aggregation**
  Central server:
  $\{w_t^1, w_t^2, \ldots, w_t^n\} \rightarrow w_{t+1}$

Variations on local update:

Type	Local update
Synchronous	fix no. of training samples
Semi-synchronous[2]	fix training time

1. Adaptive parameter freezing(APF)[1] has been introduced to synchronous FL to reduce communication overhead
2. Semi-synchronous FL reduces idle time for processors and enables faster convergence

This study extends APF to semi-synchronous FL. Several experiments were done to show the compatibility of APF with semi-synchronous FL.

Semi-synchronous FL with APF

On each client $i \in \{1, 2, \ldots, N\}$ initialize

I_{frozen} indicator of whether a parameter is frozen(1) or not(0)

T_{frozen} frozen period for a parameter

F_c stability check frequency

T threshold on parameter perturbation P_k below which to judge a parameter as being stable

Incre increment of T_{frozen} if a parameter is identified as stable

t training time for clients in one communication round

At communication round k,

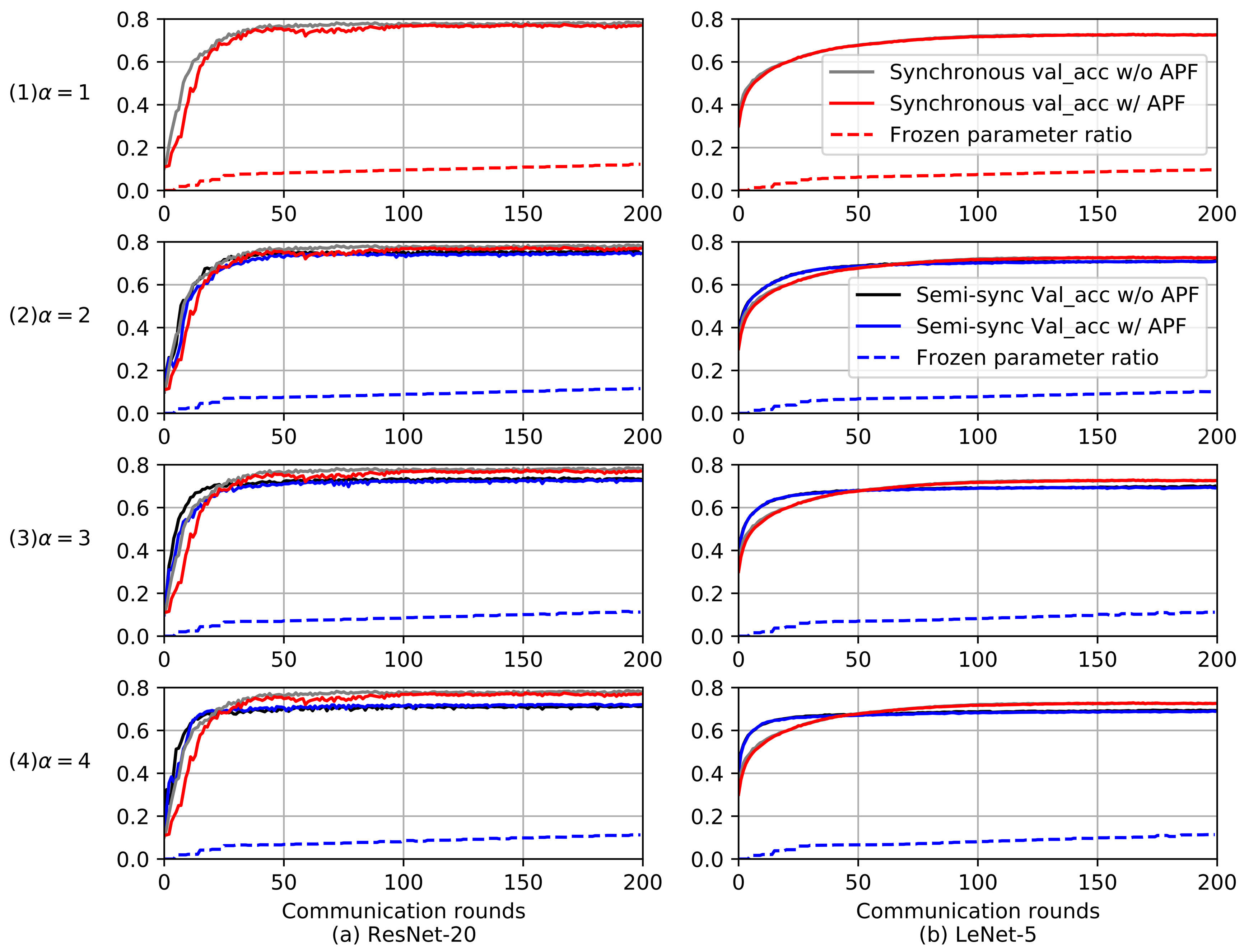
1. Clients $i \in \{1, 2, \ldots, N\}$ start to update the unfrozen parameter $\hat{w}_k$ indicated by I_{frozen} until time t is up
2. Clients $i \in \{1, 2, \ldots, N\}$ send the updated unfrozen parameters $\hat{w}_k^i$ and the number of batches used n_i to the central server
3. Central server aggregates the unfrozen parameters $\hat{w}_k^1, \ldots, \hat{w}_k^N$ to form a new partial global parameter $\hat{w}_{k+1} = \frac{\sum_{i=1}^N n_i \times \hat{w}_k^i}{\sum_{i=1}^N n_i}$
4. Central server broadcasts $\hat{w}_{k+1}$ to each client
5. Each client restores the full global parameter w_{k+1}
6. If $k \bmod F_c = 0$, then for $I_{frozen} = 0$, each client
   (a) calculate $P_k = \frac{|E_k|}{E_k^{abs}}$, where
   $E_k = \lambda E_{k-1} + (1 - \lambda)(w_{k+1} - w_k)$, and
   $E_k^{abs} = \lambda E_{k-1}^{abs} + (1 - \lambda)|w_{k+1} - w_k|$
   (b) $Incre = Incre + F_c$ if $(P_k \leq T)$ else 0
   (c) $T_{frozen} = T_{frozen} + Incre$
   (d) $I_{frozen} = (k < T_{frozen})$
7. Each client saves w_{k+1}
8. $k = k + 1$

Experiment setup

FL architecture
- 50 clients
- CIFAR-10 dataset
- (a) ResNet-20 (b) LeNet-5
- Adam optimizer with default learning rate 1e-3
- Do stability check every $T_c = 5$ communication rounds

Assumptions:
1. The clients never leave the FL process
2. The time for a client to train one epoch of data is always fixed
3. In time t (user-determined training time for clients in one communication round), 25 clients finish 1 epoch, while the rest clients finish $\alpha \in \{1, 2, 3, 4\}$ epoch(s). i.e. $\alpha = 1$ corresponds to synchronous FL.

Results and discussions



- Semi-synchronous FL achieves convergence faster but probably at the cost of model accuracy.
- APF has almost no influence to the final model accuracy, while reducing the data transmission volume (e.g. for LeNet-5 with $\alpha = 2$, each of the 50 clients is free from transmitting up to 9.23 million 64-bit floating-point numbers in the first 100 communication rounds.)

Conclusions

In this project, adaptive parameter freezing was adapted from synchronous federated learning for the semi-synchronous one. Its similar training behavior before and after the introduction of APF implies their high compatibility. However, even though semi-synchronous FL has a faster convergence rate, it trains a model with lower accuracy. A potential future direction is to train with semi-synchronous FL first and switch to synchronous FL after several communication rounds. It is expected to achieve faster convergence at the beginning and to adjust the model towards better accuracy later.

References and Acknowledgements

[1] Chen Chen, Hong Xu, Wei Wang, Baochun Li, Bo Li, Li Chen, and Gong Zhang. Communication-efficient federated learning with adaptive parameter freezing. In *IEEE International Conference on Distributed Computing Systems*, 2021.

[2] Dimitris Stripelis and José Luis Ambite. Semi-synchronous federated learning. *arXiv preprint arXiv:2102.02849*, 2021.

The author would like to thank Dr. C.H.Ngai and her PhD student for their immense support throughout the project.