A Multiple Federated Learning Services Orchestrator in Edge Computing

Minh N. H. Nguyen and Choong Seon Hong
Department of Computer Science and Engineering, Kyung Hee University
Email: {minhnhn, cshong}@khu.ac.kr

Abstract: The federated learning scheme acquires a huge amount of user equipment that participates in the learning process without revealing user data. However, in a complex scenario with multiple federated learning services, the sharing of resources among these user equipments requires an orchestrator to manage the computation, communication resources and cooperate these learning services in local learning problem quality decisions to reduce the learning time as well as energy consumptions. In this paper, we pose the optimal resource allocation, local learning quality control in the federated learning orchestrator and propose a centralized approach based on block coordinate descent to solve the multi-convex problem.

1. Introduction

Nowadays, the proliferation of high-performance mobile devices can support decentralized learning based on local dataset. In the most simple federated learning algorithm which is FedAvg[1], the authors provide a simple mechanism of averaging the local learning weights that are updated from individual UE and local dataset. Recently, many federated learning works focus on providing advanced learning algorithms to improve learning performance. As in CoCoA+ [2], the local learning of UEs is transformed into dual problems. On the other hand, in our previous work [3], [4], we propose a resource allocation problem among user equipment (UE) for a single learning service in the wireless environment regarding *computation*, *communication latency*, and *UE energy consumption*. In this paper, we extend the resource allocation framework for multiple learning services where they share CPU resources and bandwidth resources as in Fig. 1. We formulate the multi-service Federated Learning over the wireless network as a multi-convex optimization problem **MS-FEDL** for the Federated Learning Orchestrator (FLO)**.** Accordingly, FLO takes a role to manage the computation, communication resources and cooperate these learning services in local learning problem quality decisions. We exploit the problem structure to decompose and transform it into three convex sub-problems and adopt the block coordinate descent method [5] to solve these subproblems iteratively.

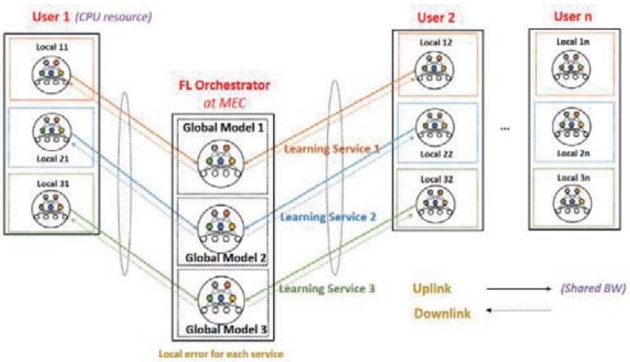


Fig. 2: Multiple Federated Learning Services Model.

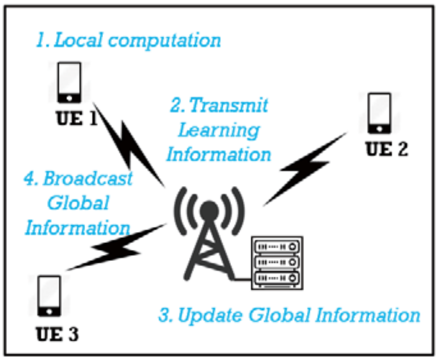


Fig. 1: General Federated Learning scheme.

Federated Learning Scheme in CoCoA+ framework [2]

- Federated Learning Loop

  S1. Local Computation: Every UE needs to solve the local learning dual problem based on primal-dual method as in CoCoA+ [2]. The output variable of this subproblem is the local changes of the dual variables.

  S2. Transmit Local Learning Information to the Federated Learning Orchestrator via a shared wireless environment

  S3. Update Global Model: The global change based on local changes are aggregated at the controller

  S4. Broadcast Global Information to all UEs: The updated global change is broadcast to all UEs.

  Until a global accuracy ϵ is achieved.

The global learning loss function as follows

$$\min_{w\in\mathbb{R}^d} L_P(w) := \frac{1}{|\mathcal{D}|}\sum_{i=1}^{|\mathcal{D}|} f_i(x_i^T w) + \lambda g(w). \qquad (1)$$

f_i is the local loss function of each UE given the input data x_i. According to the analysis of our prior work in [4] and CoCoA+ [2]:

- The normalized number of required local iterations for solving subproblem to achieve the relative accuracy θ is $log(\frac{1}{\theta})$.

- The normalized number of required global learning iteration to achieve global accuracy ϵ is $K(\theta) = (\frac{1}{1-\theta})$.

2. Multiple Federated Learning Services Framework

A. System Model

In this paper, we consider a multi-service framework of federated learning scheme with one Federated Learning Orchestrator at a BS and a set of N UEs. Each UE n stores a local dataset size $D_{s,n}$ for each federated learning service s, (i.e., local training samples $\{x_i, y_i\}_{i=1}^{D_{s,n}}$). These learning services use the shared CPU resource to compute their local learning task and bandwidth resource to transmit the updated information to Federated Learning Orchestrator.

Local computation model:

The energy consumption of each user n to solve it subproblem for service s as follows [4]

$$E_{s,n}^{cmp}(f_{s,n}) = \sum_{i=1}^{c_s D_{s,n}} \frac{\alpha_n}{2} f_{s,n}^2 = \frac{\alpha_n}{2} c_s D_{s,n} f_{s,n}^2$$

where $f_{s,n}$ is CPU-cycle frequency of the UE n, c_s *is* the number of CPU cycles for each UE to execute one sample of data belong to service s, and $\frac{\alpha_n}{2}$ is the effective capacitance coefficient of UE n's computing chipset.

Communication model:

For the communication, we consider OFDMA model with fraction w_n of total bandwidth B is allocated for each UE. The achievable transmission rate (*bps*) of UE n is defined as Shannon capacity

$$r_n^{ul}(w_n) = w_n B \log_2(1 + \frac{h_n p_n}{N_0}),$$

where p_n is the transmission power and h_n is the channel gain of UE n, and N_0 is the Gaussian noise. Then the uplink transmission time of each UE n for a service s is

$$\tau_{s,n}^{ul}(w_n) = \frac{v_s}{r_n^{ul}(w_n)},$$

where v_s is the global information size that needs to be sent to the FLO. We denote the static downlink time to broadcast the global information as τ_s^{dl}.
Thus, the communication time of a learning service s is

$$T_s^{com}(w) = \max_{n\in\mathcal{N}} \tau_{s,n}^{ul}(w_n) + \tau_s^{dl},$$

The energy consumption of uplink transmission as follows

$$E_{s,n}^{com}(w_n) = p_n \tau_{s,n}^{ul}(w_n)$$

where p_n is the transmission power of UE n.

Global Model:

Global learning time of each federated learning service s

$$T_s^{glob}(T_s^{cmp}, T_s^{com}, \theta_s) := T_s^{com} + \log(1/\theta_s) T_s^{cmp}$$

Total energy consumption of all UEs for each global iteration

$$E_s^{glob}(f_s, w, \theta_s) := \sum_{n=1}^{N} E_{s,n}^{com}(w_n) + \log(1/\theta_s) E_{s,n}^{cmp}(f_{s,n}).$$

Note that the normalized number of local iterations $log(\frac{1}{\theta_s})$ only affects to the computation time and energy.

B. Problem Formulation:

According to the sharing of CPU resource among learning services in each UE, the sharing of bandwidth resource among UEs, we propose the multi-service Federated Learning optimization problem, **MS-FEDL**, as follows

$$\begin{aligned}
\min_{f,w,\theta} \quad & \sum_{s\in\mathcal{S}} K(\theta_s)\Big(E_s^{glob}(f_s, w, \theta_s) + \kappa T_s^{glob}(T_s^{cmp}, T_s^{com}, \theta_s)\Big) \\
\text{s.t.} \quad & \sum_{s\in\mathcal{S}} f_{s,n} = f_n^{tot}, \forall n \in \mathcal{N}, (\textbf{\textit{Shared CPU}}) \\
& \sum_{n=1}^{N} w_n = 1\, (\textbf{\textit{Shared BW among users}}) \\
& f_{s,n} \geq f_{s,min}, \forall s \in \mathcal{S}, \forall n \in \mathcal{N}, \\
& w_n \geq w_{min}, \forall n \in \mathcal{N}, \\
& 0 \leq \theta_s \leq 1, \forall s \in \mathcal{S}, \\
& T_s^{cmp} \geq \frac{c_s D_{s,n}}{f_{s,n}}, \forall s \in \mathcal{S}, \forall n \in \mathcal{N}, \\
& T_s^{com} \geq \tau_{s,n}^{ul}(w_n) + \tau_s^{dl}, \forall s \in \mathcal{S}, \forall n \in \mathcal{N}.
\end{aligned}$$

where $K(\theta_s)$ is the required number of global iterations, f_n^{tot} is the total CPU frequency of UE n. The decision variables include the CPU frequency $f_{s,n}$ for each service s, the fraction w_n of total uplink bandwidth, and the relative accuracy θ_s of the local learning problem in each UE.

For the solution approach, we use block coordinate descent (BCD) in [5] to iteratively solve three following subproblems

SUB1 : **CPU Allocation**

$$\begin{aligned}
\min_{f,T^{cmp}} \quad & \sum_{s\in\mathcal{S}} \frac{\log(1/\hat{\theta}_s)}{1-\hat{\theta}_s}\Big(\sum_{n\in\mathcal{N}} \frac{\alpha_n}{2} c_s D_{s,n} f_{s,n}^2 + \kappa T_s^{cmp}\Big) \\
\text{s.t.} \quad & T_s^{cmp} \geq \frac{c_s D_{s,n}}{f_{s,n}}, \forall s \in \mathcal{S}, \forall n \in \mathcal{N}, \\
& \sum_{s\in\mathcal{S}} f_{s,n} = f_n^{tot}, \forall n \in \mathcal{N}, (\textbf{\textit{Shared CPU}}) \\
& f_{s,n} \geq f_{s,min}, \forall s \in \mathcal{S}, \forall n \in \mathcal{N}.
\end{aligned}$$

SUB2 : **BW Allocation**

$$\begin{aligned}
\min_{w,T^{com}} \quad & \sum_{s\in\mathcal{S}} \frac{1}{1-\hat{\theta}_s}\Big(\sum_{n\theta} p_n \tau_{s,n}^{ul}(w_n) + \kappa T_s^{com}\Big) \\
\text{s.t.} \quad & T_s^{com} \geq \tau_{s,n}^{ul}(w_n) + \tau_s^{dl}, \forall s \in \mathcal{S}, \forall n \in \mathcal{N}, \\
& \sum_{n=1}^{N} w_n = 1,\ (\textbf{\textit{Shared BW among users}}) \\
& w_n \geq w_{min}, \forall n \in \mathcal{N}.
\end{aligned}$$

Then the local error decision subproblem can be solved by giving the computed energy consumption of computation, communication, computation time, and communication time as $\hat{E}_s^{com}, \hat{E}_s^{cmp}, \hat{T}_s^{com}, \hat{T}_s^{cmp}$

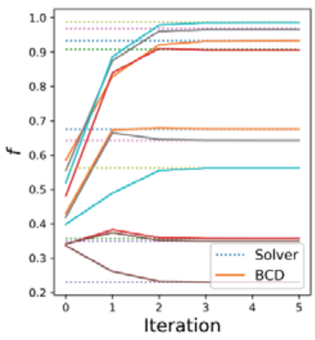
SUB3 : **Local Error Decision**

$$\begin{aligned}
\min_{\theta} \quad & \sum_{s\in\mathcal{S}} \frac{1}{1-\theta_s}\Big(\hat{E}_s^{com} + \kappa \hat{T}_s^{com} - \log(\theta_s)\big(\hat{E}_s^{cmp} + \kappa \hat{T}_s^{cmp}\big)\Big) \\
\text{s.t.} \quad & 0 \leq \theta_s \leq 1, \forall s \in \mathcal{S}.
\end{aligned}$$

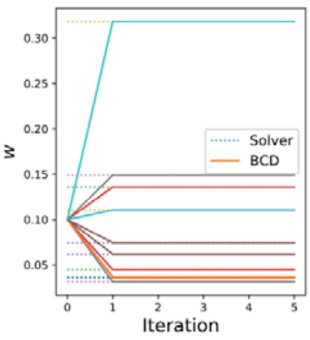
These subproblems are convex and can be solved by the solver IPOPT [6].

3. Simulation Results

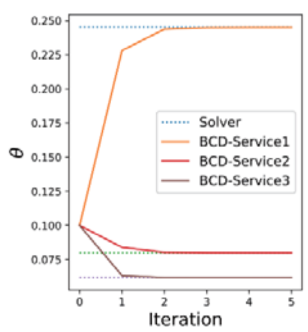
As similar setting to our prior work [4], we consider three learning services, ten heterogeneous UEs in the learning model. For the wireless communication model, the UE channel gains follow the exponential distribution with the mean $g_0(d_0/d)^4$ where $g_0 = -40$ dB and the reference distance $d_0 = 1$m. The distance between these devices and the wireless access point is uniformly distributed between 2 and 50 m. In addition, $B = 1$ MHz, $\sigma = 10^{-10}$ W. For UE computation model, we set the training size of each UE as uniform distribution in $5-10$ MB, c_n is uniformly distributed in $10-30$ cycles/bit, f_n^{max} is uniformly distributed in $1.0-2.0$ GHz, $f_n^{min} = 0.3$ GHz.



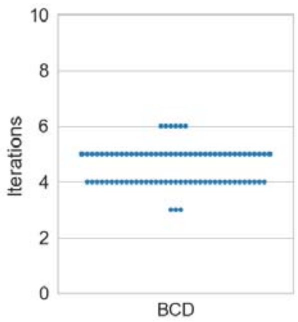
a) f convergence of service 1



b) w convergence



c) θ convergence



d) Number of iterations

Fig. 3: The convergence analysis the centralized solution.

The BCD solutions are converged to the optimal solution of the solver in Fig. 3. As in Fig. 3d, the BCD algorithm requires 5 iterations on average to achieve the convergence condition. In Fig. 3c, the service 1 has the highest error decision that means the least required number of local iterations.

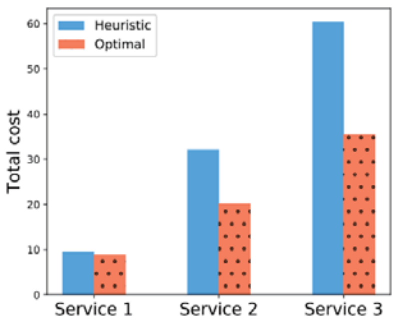


Fig. 4: Equal resource allocation vs optimal comparison.

In Fig. 4, we significantly reduce the total cost which includes learning time and energy consumption for each service compared to the heuristic approach where all of learning services receive the equal amount of CPU cycle and all users use the same amount of bandwidth. In heuristic approach, only sub3 is being solved by the controller to decide the local learning error decision for each learning service.

4. Conclusion

In this paper, we analyzed a multi-service federated learning scheme that is managed by a federated learning orchestrator. We first formulate the optimization model for computation, communication resource allocation and the local learning error decision among learning services regarding the learning time and energy consumption of UEs. We then decompose multi-convex problem into three convex sub-problems and solve the problem by using block coordinate descent algorithm. The simulation results show the convergence of the algorithm to the optimal solution. In future work, we consider extending the centralized approach to a scalable decentralized approach among services to preserve the privacy of each service.

Acknowledgement

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea goverment(MSIT) (No.2019-0-01287, Evolvable Deep Learning Model Generation Platform for Edge Computing). *Dr. CS Hong is the corresponding author.

References:

[1] McMahan, H. Brendan et al. “Communication-Efficient Learning of Deep Networks from Decentralized Data.” *Proceedings of the 20th AISTATS* (2016).

[2] C. Ma, J. Konecny, M. Jaggi, V. Smith, M. I. Jordan, P. Richtarik, and M. Takac, “Distributed Optimization with Arbitrary Local Solvers,” *Optimization Methods Software*, vol. 32, no. 4, pp. 813–848, Jul. 2017.

[3] Minh N. H. Nguyen, Choong Seon Hong, "Optimization Design Model for Decentralized Learning in Wireless Networks", 2018 년 한국컴퓨터종합학술대회(KSC 2018), 2018.12.19~12.21.

[4] Nguyen H. Tran, Wei Bao, Albert Zomaya , Minh N.H. Nguyen and Choong Seon Hong, “Federated Learning over Wireless Networks: Optimization Model Design and Analysis,” *IEEE International Conference on Computer Communications (INFOCOM 2019)*, April 29 - May 2, 2019, Paris, France.

[5] Y. Xu and W. Yin, “A Block Coordinate Descent Method for Regularized Multiconvex Optimization with Applications to Nonnegative Tensor Factorization and Completion,” *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1758–1789, jan 2013.

[6] A. Wachter and L. T. Biegler, “On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming,” Mathematical Programming, vol. 106, no. 1, pp. 25–57, Mar. 2006.