Optimization over Client Selection in Efficient Federated Learning

Dixi Yao
dixi.yao@mail.utoronto.ca
Student ID: 1009018313

Kai Wang
kaikai.wang@mail.utoronto.ca
Student ID: 1008111945

***Abstract*— Federated learning emerges as a branch of machine learning for solving problems by multiple clients in a decentralized fashion, where various clients learn the local models on the local datasets, and the master server aggregates the updates from selected clients for global updates. Therefore, efficiently selecting clients for participating in global aggregation is very crucial when considering the trade-off between performance and communication efficiency in federated learning. In this project, we formulate the client selection problem as solving a convex optimization problem to obtain the optimal selection solution. Experimental results on the benchmark dataset show that our problem convex optimization based client selection method achieves the same performance to FedAvg baseline model while taking significantly reduced time, improving the communication efficiency in federated learning.**

I. Introduction

Federated learning [1], [2] is an emerging sub-area of distributed optimization, which instructs various clients to jointly learn a global model without data sharing as shown in Fig. 1. Unlike traditional distributed systems, in federated learning, because the connection among clients can be unstable and pose privacy concerns, only a subset of clients are selected in each communication round. One of the bottleneck issues is communication, since the data transmission rates of internet connections vary from different end-users. Due to network capacity, the server may constrain the number of communicated clients within one round. In synchronous federated learning, we assume all clients communicate with the server simutinously and they work parallelly. As a result, the order of how to select clients does not matter but a cutting-edge problem in federated learning is how to wisely select the subset of clients so as to achieve efficiency, or we can call it biased client participation. It is important to analyze and understand biased client selection strategies because they can sharply boost communication efficiency in heterogeneous environments by preferentially selecting clients [3].

To efficiently reduce server-client communications, the server can sample a subset of clients participating in communication by various principles. In [1], authors proposed a FedAvg algorithm to uniformly select clients and replaced the contribution of non-sampled clients by current global model. However, the bias will be introduced into the generated model by FedAvg since it is different from the deterministic aggregation of each client. To tackle this issue, authors of [4] proposed unbiased sampling approach to consider the new global model as the average contribution of the sampled clients, where the sampling performs on multinomial distribution whose clients probabilities correspond to their related sample size. Researchers of [5] proposed an optimal sampling scheme where all clients are permitted to participate in the computation of updates. However, only some important updates are selected to communicate back to the server. Authors of [6] proposed a FedCS protocol to mitigate the client selection problem according to resource constraints, which lets the server aggregate as many client updates as possible for performance improvement. [7] introduced a clustering sampling for clients selection to solve the biased and non-optimal problems, which reduces the variance of the weights aggregated by clients and improves the representativity of clients.

In this work, we formulate the client selection strategy as a convex optimization problem to improve communication efficiency by obtaining an optimal selection of participated clients. By comparing with the FedAvg algorithm on the same dataset, our proposed convex optimization method achieves a similar performance while involving less time, which indicates the efficiency of proposed client selection method on reducing communication cost between clients and server.

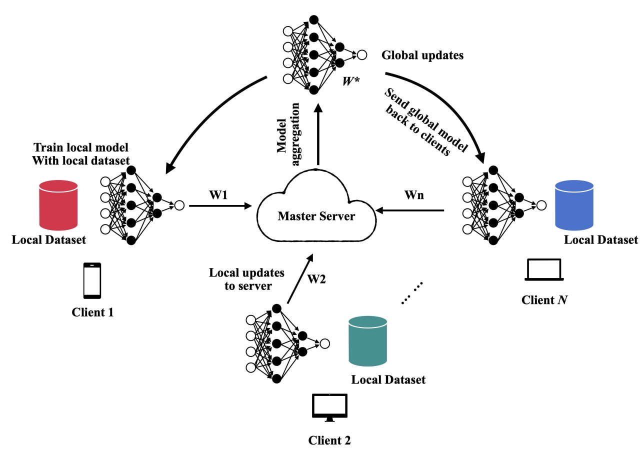

Fig. 1: Overview of client-server architecture in Federated Learning [8]

II. Problem Formulation

In this section, we first give out the formal formulation of the client selection problem, and then describe how to convert it into a convex optimization problem.

A. Problem Statement

In this project, the client selection problem we hope to solve can be written as follows, a simple version of the formulation:

$$\begin{aligned} \text{minimize} \quad & \mathbb{E}[T_{total}(\mathbf{q})] \\ \text{subject to} \quad & \mathbb{E}[F(\mathbf{w}^R\mathbf{q}] - F^* \leq \epsilon \\ & \sum_1^n q_i = 1 \\ & q_i > 0, \forall i \in [n], \end{aligned} \tag{1}$$

where $\mathbf{q} \in \mathbf{R}^n$ are the probabilities of selecting clients (only consider i.i.d. manner) in each communication round, T_{total} is the total learning time related to variable q, which we will later solve out. Especially, in the constraint, we also add the requirement that after R communication rounds of federated learning, the error between expectation of trained models $F(\mathbf{w}^R\mathbf{q})$ and the optimal value F^* should be less than a given small number ϵ.

B. Convex Formulation

In this section, we will show how to turn equation 1 (which contains abstract formulations) into an convex optimization problem. With given n clients, and assume objective function of each client: f_i is lipschitz smooth and ∇f_i is upper bounded by G_i. To deviate the formulation of $\mathbf{E}[T_{total}]$, we consider two candidates: the total communication round: R and averaged time spent in each communication round: t.

For communication round R, first of all, as $\mathbb{E}[T_{total}(\mathbf{q})]$ decreases along the training process, and we want to minimize the total training time. The system can stop as long as the equation $\mathbb{E}[F(\mathbf{w}^R\mathbf{q}] - F^* = \epsilon$ meets. Then, we can use the theory of convergence rate in Federated Learning [9] to have such formulation

$$\mathbb{E}[F(\mathbf{w}^R\mathbf{q}] - F^* = \epsilon \leq \frac{1}{R}\left(\alpha \sum_{i\in[n]} \frac{p_i^2 G_i^2}{q_i} + \beta\right) \tag{2}$$

where ϵ is a small error and α is a constant decided by system environment, representing global convergence rate in federated learning. p_i means the promotion of samples on each client regarding to all samples, β is also a constant related to the actual system setting, representing minimal gradients needed to be changed to reach the best accuracy.

This formulation shows the relationship between the total communication rounds we need and the error we achieve. In federated learning, we want to give each client a different selected probability so as to minimize the overall time regarding achieving a given small error. Then, we can have

$$\mathbb{E}[T_{total}] = \frac{t}{\epsilon}\left(\alpha \sum_{i\in[n]} \frac{p_i^2 G_i^2}{q_i} + \beta\right) \tag{3}$$

Regarding the averaged time spent in each communication round, since we only consider the synchronous setting here, we can use t as a constraint to define the maximum time of local training of clients in each communication round. To write this in mathematical formulation, it means

$$\gamma^T \mathbf{q} \leq t, \tag{4}$$

where $\boldsymbol{\gamma} \in \mathbf{R}^n = (\gamma_1, \cdots, \gamma_n)$, meaning the local computation time including communication and training time on each client and this is completely decided by the system settings.

Substitute equation 2, equation 3, and equation 4 into equation 1, our problem is

$$\begin{aligned} \text{minimize} \quad & t\left(\frac{\alpha}{\beta}\sum_{i\in[n]} \frac{p_i^2 G_i^2}{q_i} + 1\right) \\ \text{subject to} \quad & \mathbf{1}^T\mathbf{q} \leq 1 \\ & \mathbf{1}^T\mathbf{q} \geq 1 \\ & \boldsymbol{\gamma}^T\mathbf{q} \leq t \\ & \mathbf{q} \succeq \mathbf{0}, \end{aligned} \tag{5}$$

where our variable is $\mathbf{q} \in \mathbf{R}^n_{++}$. And $t \in \mathbf{R}_{++}$, $\mathbf{p} \in \mathbf{R}^n_{++}$, $\boldsymbol{\gamma} \in \mathbf{R}^n_{++}$, $\mathbf{G} \in \mathbf{R}^n_{++}$, $\alpha \in \mathbf{R}_{++}$ and $\beta \in \mathbf{R}$ are parameters given by the system which related to the distribution of data, the computation ability of system and local optimization targets, where we hope to choose the best policy of selecting clients instead of random selection to achieve the best overall time. This problem is similar to the portfolio distribution problem introduced in the lecture, and here the proportion of stock investment is changed into the probability of selecting clients and in the objective, we have item of inverse of q_i.

C. Analysis of Convexity

The domain of the objective function is $\mathbf{R}^n_{++}$. The objective function is a non-negative combination of convex function $F(x) = \frac{1}{x}, x > 0$, which is a convex function. All inequality constraints are affine. So, this is a convex optimization problem. To step further, this is a geometric programming (GP) problem, as the objective function and inequality functions are all polynomials.

D. Obtain $\mathbf{G}$ and $\frac{\alpha}{\beta}$

First, we can obtain $\mathbf{G}$ through running the baseline method of federated averaging. Then following the instruction in paper [9], we can use two sampling distribution q_1 which is the uniform distribution and q_2, where $q_{2,i} = p_i$ and calculate the R_1 and R_2 using the baseline method of federated averaging, then derive $\frac{\alpha}{\beta}$ with this formulation

$$\frac{R_1}{R_2} = \frac{\frac{\alpha}{\beta} n \sum p_i^2 G_i^2 + 1}{\frac{\alpha}{\beta}\sum p_i G_i^2 + 1}. \tag{6}$$

III. Experiments

A. Experimental Setups

We evaluated our proposed convex optimization method on the FEMNIST dataset, which is used for image classification tasks with federated settings. This dataset totally includes

3,550 clients while each one has different portions of samples, leading to 80,5263 samples. To obtain the required parameters in convex problem mentioned above, we first implement FedAvg algorithm as baseline to generate necessary parameters including $\mathbf{p}$, $\boldsymbol{\gamma}$, $\mathbf{G}$, α and β since they are only related to dataset itself. To simplify the experiment procedure, only 100 clients were selected to participate in the global update on the server. The target classification accuracy is set as 70% to control the experiment termination, which means all experiments will stop until the target accuracy is achieved. Due to resource and time limitations, we did not repeat experiments, but the random seed is completely fixed across all experiments.

B. Client Selection Strategies

In this work, three different client selection strategies were used for federated learning, including proposed convex methods as well as FedAvg methods with two different configurations as shown in Fig. 2. In the FedAvg, the sampling rate of clients $\mathbf{q}$ is all the same uniform distribution. For FedAvg with equivalent $\mathbf{q}$ and $\mathbf{p}$, the client selection is consistent with dataset distribution. Different from that, our proposed selection approach is based on the optimal solution solved by the convex optimization problem.

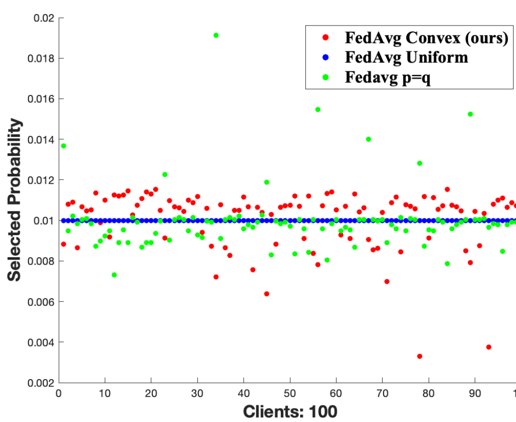

Fig. 2: Different client selection methods in 100 clients

C. Experimental Results

We compare proposed client selection methods with the ones used in FedAvg on image classification performance as shown in Fig. 3. To have a fair comparison, except for the sample rate of selected clients, other configurations are maintained to be the same, such as network structure and training procedures. From Fig. 3, we can observe that the classification performance of the three selection methods is accordingly improved as time increases. More specifically, compared to FedAvg baseline model, both our proposed method and Fedavg with $\mathbf{q}$ equal to $\mathbf{p}$ achieve the target accuracy performance in advance. Moreover, proposed convex optimization method takes least time to obtain the same target accuracy while having the better convergence rate, which indicates that proposed client selection method is more efficient in decreasing communication time of federated learning.

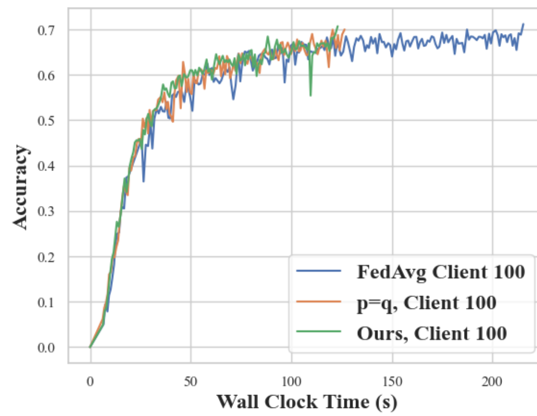

Fig. 3: Comparison results among different client selection methods

In addition, TABLE I compares the required time to achieve the target accuracy among different client selection methods. As shown in Table I, Fedaveg baseline needs about 215.31 seconds to arrive to 70% target accuracy, while our proposed convex-based selection method only requires to take about half time (123.07 seconds) to obtain the same classification performance. The possible reason is that FedAvg baseline adopted the uniformly distributed selection method to make each client have the same participated possibility. However, our proposed convex-based method efficiently selects clients that contribute more accuracy performance with less communication time with master server by solving a convex optimization problem.

TABLE I: Time to the target accuracy among different methods

Method	Time (s)	Rounds
FedAvg Baseline	215.31	201
FedAvg p = q	126.49	134
FedAvg Convex (ours)	**123.07**	134

IV. Conclusion

In this project, we formulated the client selection problem in federated learning into a convex optimization problem, and obtained an optimal selection strategy for clients participating in the global aggregation at server. Experimental results indicate that our proposed convex optimization method achieved the same performance as FedAvg baseline with significantly reduced time.

References

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.

[2] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings *et al.*, "Advances and open problems in federated learning," *Foundations and Trends® in Machine Learning*, vol. 14, no. 1–2, pp. 1–210, 2021.

[3] Y. J. Cho, J. Wang, and G. Joshi, "Client selection in federated learning: Convergence analysis and power-of-choice selection strategies," *arXiv preprint arXiv:2010.01243*, 2020.

[4] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *Proceedings of Machine Learning and Systems*, vol. 2, pp. 429–450, 2020.

[5] W. Chen, S. Horvath, and P. Richtarik, "Optimal client sampling for federated learning," *arXiv preprint arXiv:2010.13723*, 2020.

[6] T. Nishio and R. Yonetani, "Client selection for federated learning with heterogeneous resources in mobile edge," in *ICC 2019-2019 IEEE international conference on communications (ICC)*. IEEE, 2019, pp. 1–7.

[7] Y. Fraboni, R. Vidal, L. Kameni, and M. Lorenzi, "Clustered sampling: Low-variance and improved representativity for clients selection in federated learning," in *International Conference on Machine Learning*. PMLR, 2021, pp. 3407–3416.

[8] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, and Y. Gao, "A survey on federated learning," *Knowledge-Based Systems*, vol. 216, p. 106775, 2021.

[9] B. Luo, W. Xiao, S. Wang, J. Huang, and L. Tassiulas, "Tackling system and statistical heterogeneity for federated learning with adaptive client sampling," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*. IEEE, 2022, pp. 1739–1748.