Multi-Criterion Client Selection for Efficient Federated Learning

Mehreen Tahir[1], **Muhammad Intizar Ali**[2]

[1]SFI Centre for Research Training in Machine Learning, Dublin City University, Dublin, Ireland
[2]School of Electronic and Computer Engineering, Dublin City University, Dublin, Ireland
mehreen.tahir2@mail.dcu.ie, ali.intizar@dcu.ie

Abstract

Federated Learning (FL) has received tremendous attention as a decentralized machine learning (ML) framework that allows distributed data owners to collaboratively train a global model without sharing raw data. Since FL trains the model directly on edge devices, the heterogeneity of participating clients in terms of data distribution, hardware capabilities and network connectivity can significantly impact the overall performance of FL systems. Optimizing for model accuracy could extend the training time due to the diverse and resource-constrained nature of edge devices while minimizing training time could compromise the model's accuracy. Effective client selection thus becomes crucial to ensure that the training process is not only efficient but also capitalizes on the diverse data and computational capabilities of different devices. To this end, we propose FedPROM, a novel framework that tackles client selection in FL as a multi-criteria optimization problem. By leveraging the PROMETHEE method, FedPROM ranks clients based on their suitability for a given FL task, considering multiple criteria such as system resources, network conditions, and data quality. This approach allows FedPROM to dynamically select the most appropriate set of clients for each learning round, optimizing both model accuracy and training efficiency. Our evaluations on diverse datasets demonstrate that FedPROM outperforms several state-of-the-art FL client selection protocols in terms of convergence speed, and accuracy, highlighting the framework's effectiveness and the importance of multi-criteria client selection in FL.

Introduction

The shift from centralized data centers to edge computing has highlighted Federated Learning (FL) as a key approach for efficient, privacy-preserving model training on edge devices (McMahan and Ramage 2017; Wang et al. 2019b). This paradigm shift, however, introduces challenges due to the inherent heterogeneity in edge devices' hardware, communication, and data resources (Bonawitz et al. 2019; Hsieh et al. 2020; Zawad et al. 2021), directly impacting model accuracy and system efficiency. Current FL protocols often overlook this heterogeneity, focusing primarily on either optimizing statistical model performance (Li et al. 2020; Zhang et al. 2020) or system efficiency (Brendan et al. 2016), but not both. Our research (Tahir and Ali 2022) indicates that incorporating updates from devices with low-quality data or delays can adversely affect the overall system performance. Therefore, a strategic client selection strategy, considering the constraints of edge devices, is essential for high-quality federated solutions.

Existing client selection methods often simplify the issue, focusing on singular criteria like computational resources (Nishio and Yonetani 2019; Sasindran, Yelchuri, and Prabhakar 2023; Chen et al. 2024) or data quality (Cho, Wang, and Joshi 2020; Lin et al. 2022; Deng et al. 2022), potentially leading to biased model updates. To address these challenges, we propose FedPROM, a multi-criterion client selection framework that aims to jointly optimize statistical model utility and system efficiency. We formulate the client selection problem as a multi-criterion-decision-making (MCDM) problem and employ the Preference Ranking Organization METHod for Enrichment of Evaluations (PROMETHEE) (Brans and Vincke 1985) to systematically evaluate the clients based on multiple criteria such as availability, computational resources (e.g., CPU, memory, storage, power), network conditions (e.g., bandwidth, latency) and data quality (e.g., data size, accuracy) as well as practical constraints like budget limits (B). FedPROM's comprehensive approach enables a balanced and effective ranking of clients, ensuring an optimized trade-off between accuracy and training time. Our results demonstrate how a multi-criterion framework can significantly enhance both the efficiency and effectiveness of FL systems, offering valuable insights for future advancements in this domain.

Related Work

Client selection in FL has been the subject of extensive research in recent years due to its critical role in determining the efficiency and performance of FL systems with selection strategies ranging from simple random selection to advanced heuristics (Nishio and Yonetani 2019), machine learning-based approaches (Zhang, Lin, and Zhang 2022), and those based on data quality measures (Sattler et al. 2019; Goetz et al. 2019; Cho, Wang, and Joshi 2020). Despite their diversity, the effectiveness of these strategies in balancing efficiency and performance in FL environments remains a topic of active debate. For instance, random selection (Brendan et al. 2016) ensures fairness but may neglect client hetero-

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

geneity, leading to inefficiencies. Heuristic methods, while accounting for aspects such as computational power, lack adaptability and lead to biased outcomes (Lim et al. 2020). Similarly, methods focusing on local client loss (Goetz et al. 2019; Cho, Wang, and Joshi 2020) and client contribution (Lin et al. 2022; Pandey, Nguyen, and Popovski 2022), often disproportionately favor high-performing clients and overlook hardware resource constraints.

Other client selection techniques such as those based on reinforcement learning and cluster-based selection, offer dynamic and adaptive client selection (Wang et al. 2019a; Cheng et al. 2022; Deng et al. 2022; Albelaihi et al. 2023). However, their utility is counterbalanced by the need for substantial meta-data and computational resources, increasing the system's complexity and computational overhead.

Recognizing these challenges, this paper introduces FedPROM, a novel framework designed to systematically consider hardware capabilities, network conditions, and data quality of edge devices. This unified approach aims to effectively manage the trade-off between performance and efficiency, proposing an optimal client selection method for efficient and effective FL.

FedPROM: A Multi-Criterion Client Selection Protocol For Federated Learning

Problem Definition

Given the inherent heterogeneity of the edge devices in terms of varied computation and communication capacities and data distribution, the challenge lies in selecting an optimal subset of clients ($S \subseteq C$) that ensures high performance and swift convergence for the global model. This selection must balance accuracy and training efficiency, often at odds due to trade-offs involved.

Formally, let $C = \{c_1, c_2, ..., c_n\}$ represent the set of all available clients. Each client c_i is characterized by an attribute vector $V_i = [H_i, N_i, Q_i]$, a local dataset D_i of size $|D_i|$ and an associated cost $cost(c_i)$. Here, $H = \{h_1, h_2, ...\}$ represent the client's system-level metrics (where h_i represents available CPU, memory, storage, battery level, etc.), $N = \{n_1, n_2, ...\}$ denotes network-level attributes (like bandwidth and reliability), and Q represents the quality of the client's data determined by an initial measure of local accuracy (gauged against the publisher's test set) and the size of the client's dataset $|D_i|$. The primary objective is to identify a $S \subseteq C$, such that the global model achieves maximum accuracy in the shortest possible time when trained using data from the clients in S, while the cumulative cost remains within a predefined budget B.

Problem Formulation

In this study, we approach client selection in FL as a dual-objective optimization problem. From a pool of clients C, our goal is to identify a subset $S \subseteq C$ that optimally contributes to the global model's performance, under a set budget constraint. This problem can be formulated as follows:

$$\begin{cases} \max_{S \subseteq \mathcal{C}} & A(S) = f_A(S) \\ \min_{S \subseteq \mathcal{C}} & T_{\text{total}}(S) = f_T(S) \end{cases}$$

Subject to:

$$\sum_{c_i \in S} \text{cost}(c_i) \leq B,$$

$$S \subseteq \mathcal{C}.$$

Where:

- $A(S)$ represents the aggregated accuracy of the global model using the selected subset of clients S. The function $f_A(S)$ calculates this accuracy, factoring in the individual contributions of clients based on their characteristics and the aggregation method used. [1]
- $T_{total}(S)$ denotes the total training time for the global model using the selected subset S. This function aggregates the individual training times of each client in S.
- The budget constraint ensures the total cost of the selected clients does not exceed the available budget B.

Determining the optimal subset S that satisfies these objectives is computationally challenging. In fact, the problem can be shown to be NP-hard, as it involves combinatorial optimization over the set of all possible client subsets.

Theorem: The multi-criterion optimization problem for client selection in FL, as we formulated above, is NP-hard.

Proof: We demonstrate the NP-hardness of our client selection problem by constructing a reduction from the Knapsack problem, which is a well-known NP-hard problem.

In the Knapsack problem, the goal is to select items to maximize the total value without exceeding a weight limit. We map this problem to our client selection context as follows:

- *Items in Knapsack*: Each item in the Knapsack problem corresponds to a client in the FL network.
- *Weight of Items*: The weight of each item in the Knapsack problem is analogous to the cost associated with selecting a client for training.
- *Value of Items*: The value of an item in the Knapsack problem maps to a combined metric in the FL problem, representing each client's contribution to model accuracy and the efficiency of training. This metric could be a function of the client's data quality and the reciprocal of its expected training time, where shorter training times are more valuable.
- *Weight Capacity*: The weight capacity in the Knapsack problem equates to the total available budget in our problem, encompassing overall resource limitations.

The transformation function T takes an instance of the Knapsack problem (items, weights, values, and capacity) and maps it to an instance of the client selection problem with corresponding clients, resource costs, contributions to accuracy, and training time efficiency under a budget constraint.

[1]The function $f_A(S)$, represents the expected accuracy of the global model for a subset of clients, based on the aggregation method used. For illustrative purposes, consider the FedAvg aggregation mechanism, where $f_A(S)$ translates to the mean of the model updates provided by the selected clients. In practice, $f_A(S)$ can represent any aggregation mechanism.

Since solving the transformed problem effectively solves the Knapsack problem instance and the Knapsack problem is NP-hard, it follows that the client selection problem, with its dual objectives of maximizing accuracy and minimizing training time under budget constraints, is also NP-hard. This complexity arises from the need to optimize multiple conflicting criteria simultaneously.

Optimal Multi-Criterion Client Selection Using FedPROM

To address the NP-hard multi-criterion client selection problem in FL, we introduce FedPROM. This approach transforms the complex problem into a manageable ranking task by systematically evaluating clients based on multiple attributes, including system-level metrics H, network attributes N, and data quality Q, normalized for comparability. Furthermore, utilizing the PROMETHEE decision-making protocol enables FedPROM to effectively navigate the trade-off between multiple objectives.

Each client c_i is assessed through pairwise preference functions. For two clients c_i and c_k, the preference function for a criterion j is defined as follows:

$$P_j(d_{ij,ik}) = \begin{cases} 0 & \text{if } d_{ij,ik} \leq q_j, \\ \frac{d_{ij,ik} - q_j}{p_j - q_j} & \text{if } d_{ij,ik} > q_j, \end{cases} \tag{1}$$

where $d_{ij,ik} = |g_j(c_i) - g_j(c_k)|$ is the absolute difference in the j-th criterion between clients c_i and c_k, q_j is the indifference threshold, and p_j is the preference threshold for the j-th criterion.

The global preference score $\pi(c_i, c_k)$ for client c_i over client c_k is given by:

$$\pi(c_i, c_k) = \sum_{j=1}^{m} w_j P_j(d_{ij,ik}), \tag{2}$$

where w_j is the weight for the j-th criterion, and m is the total number of criteria. The outranking flows ϕ^+ and ϕ^- are defined as the average of global preference scores, indicating the extent to which a client is preferred over or less preferred than other clients, respectively:

$$\phi^+(c_i) = \frac{1}{n-1} \sum_{k \neq i} \pi(c_i, c_k), \tag{3}$$

$$\phi^-(c_i) = \frac{1}{n-1} \sum_{k \neq i} \pi(c_k, c_i), \tag{4}$$

where n is the number of clients considered. The final ranking is determined by the net outranking flow $\phi(c_i)$:

$$\phi(c_i) = \phi^+(c_i) - \phi^-(c_i), \tag{5}$$

Clients with higher net flows are deemed optimal for selection, constrained by the budget B. This approach reduces the complexity from exponential in a brute-force scenario to polynomial, making it feasible for large-scale FL setups.

Experimental Setup

To assess the effectiveness of the FedPROM client selection protocol, we conducted experiments using the Fashion-MNIST and CIFAR-10 datasets. These datasets were non-IID partitioned across 100 simulated clients, each receiving a randomized allocation of 50 to 1000 images to emulate real-world data distribution. The client devices varied in capability and were categorized as low-end, mid-range, and high-end. The simulation environment ran on a workstation with a 12th Gen Intel(R) Core(TM) i7-12700K CPU clocking at 3.60 GHz, and 64GB RAM to mimic an FL environment. Network conditions, such as bandwidth and latency, were also modeled to reflect varying connectivity qualities. A latency factor was also integrated into the simulation based on the device and network capabilities of each client's device to closely reflect the real-world FL environment.

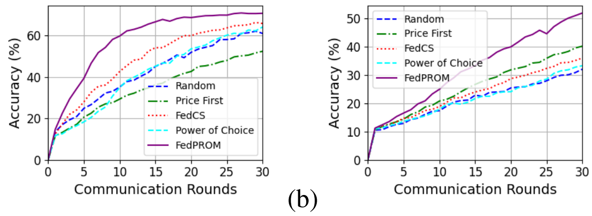
For model training, Fashion-MNIST utilized a custom convolutional neural network, with two convolutional layers, a dropout layer, and two fully connected layers. For the CIFAR dataset, we used a modified version of the pre-trained ResNet-50 model, adjusted with an adaptive average pooling layer and a final fully connected layer outputting ten classes. Both datasets underwent 30 global training rounds, with a learning rate of 0.01, to compare the performance of FedPROM against other client selection strategies.

Evaluation and Results

Our evaluation benchmarks the FedPROM client selection strategy against FedAvg and advanced strategies across multiple metrics. The comparison includes:

- **Random Selection**: Clients are selected completely at random, providing a baseline that ensures no bias in selection but lacks optimization for learning efficiency (McMahan and Ramage 2017).
- **Adapted FedCS (Resource-based)**: Selection is based on the clients' available computational and communication resources, aiming to include the most reliable and efficient participants (Nishio and Yonetani 2019).
- **Price First**: Prioritizes clients offering the least expensive computational resources, focusing on cost minimization for federated learning operations.
- **Power of Choice (pow-d)**: Employs a strategy based on local client loss, selecting clients that are likely to offer the greatest immediate improvement to model accuracy (Cho, Wang, and Joshi 2020).

Furthermore, FedPROM, our proposed method, is assessed for its efficiency and effectiveness in enhancing federated learning outcomes. We evaluated these strategies based on accuracy progression over training rounds, Time of Arrival (ToA) at designated accuracy levels, and performance post 30 rounds. The training involved 100 clients with varying participation levels (10%, 20%, and 50% clients) and varying budget constraints.



(a) (b)

Figure 1: Performance comparison of (a) Fashion-MNIST (b) CIFAR 10 datasets with 10% participating clients.

Performance Evaluation and Discussion

Emulating the client characteristics as detailed in Section , we trained models selecting a subset of clients (10% per iteration) and recorded their accuracy progression, as depicted in Figure 1. Our analysis reveals that FedPROM consistently surpasses other strategies across varied tasks and client scales. Notably, in a CIFAR-10 task after 30 rounds, FedPROM attained a 51.98% accuracy rate, significantly outperforming the Random, Price First, FedCS, and Power of Choice strategies by 62%, 29%, 44%, and 56%, respectively. Additionally, FedPROM's robustness was evaluated with expanded client pools of 20% and 50% participation, demonstrating significant superiority over other benchmarks even at scale. While alternative protocols displayed consistent yet fluctuating performance, suggesting context-dependent effectiveness, FedPROM maintained notable stability.

Time of Arrival at a Desired Accuracy (ToA@x): We observed the changes in the accuracy on testing datasets over time and analyzed when accuracy reached a certain level. Specifically, we report Time of Arrival at a Desired Accuracy (ToA@x) and Rounds of Arrival (RoA@x) for Fashion-MNIST under non-IID data settings for 50% and 65% accuracy levels. We summarize these results in Table 1.

We can observe that FedPROM consistently outperforms the existing client selection strategies in both ToA and RoA. Specifically, for achieving a 50% accuracy level, FedPROM required only 245 seconds and 9 communication rounds marking a 52% improvement over Random, 71% over Price First, and 41% over Power of Choice in time, and 25% improvement over Random, 44% over Price First, and 31% over Power of Choice in communication rounds. For the CIFAR-10 dataset, FedPROM was the only strategy to surpass the 50% accuracy mark, thereby highlighting its superior efficiency and effectiveness.

Method	Fashion-MNIST			
	50% accuracy		65% accuracy	
	ToA	RoA	ToA	RoA
Random	515.89	12	NaN	NaN
Price First	842.35	13	1714.46	28
FedCS	**206.07**	15	NaN	NaN
pow-d	415.07	13	886.47	25
FedPROM	245	**9**	**421.61**	**20**

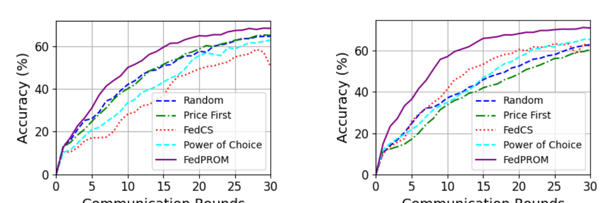
Table 1: Results obtained for Fashion-MNIST with non-IID data setting. ToA@x represents the time (in seconds) and RoA@x represents the communication rounds required to arrive at a testing classification accuracy of x (the earlier the better). NaN means that the method did not achieve the testing classification accuracy of x% in the given 30 rounds of training.

Impact of Budget: We comprehensively evaluated the impact of budget constraints on model accuracy, with budget parameters ranging from 5 to 50. Our findings, using the Fashion-MNIST dataset as a representative sample, suggest that FedPROM consistently outperforms other strategies across varying budget levels. For example, at a minimum budget of 5, FedPROM achieves an accuracy of 65.5%. This represents an 8.5% improvement over the next best-performing method, which achieves an accuracy of 59.71% under the same budget constraints. Although other methods show incremental improvements in accuracy as the budget increases, FedPROM maintains a stable high-performance level. Specifically, its accuracy hovers around the 71%-72% range when the budget is 15 or higher, indicating that FedPROM can achieve near-optimal performance without requiring additional budget allocation.

Conclusion and Future Work

This study introduces FedPROM, a novel multi-criteria client selection strategy in FL that aims to jointly optimize the performance and efficiency of the system. The empirical results from the Fashion-MNIST and CIFAR-10 datasets indicate a significant performance improvement over traditional strategies. By comprehensively evaluating clients' hardware capabilities, network conditions, and data quality, our approach boosts the overall system performance, thereby addressing a fundamental challenge in the field of FL.

Our future work aims to validate the approach with larger and more complex datasets, test its scalability, and examine its adaptability to different learning tasks. Further research also concentrates on developing dynamic client selection strategies that allow a responsive approach to the dynamic conditions of real-world FL scenarios. Ultimately, this research marks a significant step towards refining FL client selection, with the potential to substantially advance distributed machine learning practices.



(a) (b)

Figure 2: Performance comparison of Fashion-MNIST dataset for varying budget constraint (a) Budget=10 (b) Budget=20

Acknowledgments

This work is funded by Science Foundation Ireland through the SFI Centre for Research Training in Machine Learning under Grant Number 18/CRT/6183 and also under Grant Number SFI/12/RC/2289_P2 (Insight).

References

Albelaihi, R.; Alasandagutti, A.; Yu, L.; Yao, J.; and Sun, X. 2023. Deep Reinforcement Learning Assisted Client Selection in Non-orthogonal Multiple Access based Federated Learning. *IEEE Internet of Things Journal*, 1–1.

Bonawitz, K.; Eichner, H.; Grieskamp, W.; Huba, D.; Ingerman, A.; Ivanov, V.; Kiddon, C.; Konečný, J.; Mazzocchi, S.; Brendan, M. H.; Overveldt, V.; Petrou, D.; Ramage, D.; and Roselander, J. 2019. Towards Federated Learning at Scale: System Design.

Brans, J. P.; and Vincke, P. 1985. A Preference Ranking Organisation Method: (The PROMETHEE Method for Multiple Criteria Decision-Making). *Management Science*, 31: 647–656.

Brendan, M. H.; Moore, E.; Ramage, D.; Hampson, S.; and Arcas, B. A. y. 2016. Communication-Efficient Learning of Deep Networks from Decentralized Data.

Chen, S.; Tavallaie, O.; Hambali, M. H.; Zandavi, S. M.; Haddadi, H.; Lane, N.; Guo, S.; and Zomaya, A. Y. 2024. Optimization of Federated Learning's Client Selection for Non-IID Data Based on Grey Relational Analysis. arXiv:2310.08147.

Cheng, Z.; Fan, X.; Liwang, M.; Chen, N.; and Wang, X. 2022. Learning-Based Client Selection for Federated Learning Services Over Wireless Networks with Constrained Monetary Budgets. arXiv:2208.04322.

Cho, Y. J.; Wang, J.; and Joshi, G. 2020. Client Selection in Federated Learning: Convergence Analysis and Power-of-Choice Selection Strategies. *arXiv:2010.01243*, v1.

Deng, Y.; Lyu, F.; Ren, J.; Wu, H.; Zhou, Y.; Zhang, Y.; and Shen, X. 2022. AUCTION: Automated and Quality-Aware Client Selection Framework for Efficient Federated Learning. *IEEE Transactions on Parallel and Distributed Systems*, 33(8): 1996–2009.

Goetz, J.; Malik, K.; Bui, D.; Moon, S.; Liu, H.; and Kumar, A. 2019. Active Federated Learning. arXiv:1909.12641.

Hsieh, K.; Phanishayee, A.; Mutlu, O.; and Gibbons, P. B. 2020. The Non-IID Data Quagmire of Decentralized Machine Learning. arXiv:1910.00189.

Li, T.; Sahu, A. K.; Zaheer, M.; Sanjabi, M.; Talwalkar, A.; and Smith, V. 2020. Federated Optimization in Heterogeneous Networks. arXiv:1812.06127.

Lim, W. Y. B.; Luong, N. C.; Hoang, D. T.; Jiao, Y.; Liang, Y.-C.; Yang, Q.; Niyato, D.; and Miao, C. 2020. Federated Learning in Mobile Edge Networks: A Comprehensive Survey. *IEEE Communications Surveys Tutorials*, 22(3): 2031–2063.

Lin, W.; Xu, Y.; Liu, B.; Li, D.; Huang, T.; and Shi, F. 2022. Contribution-based Federated Learning client selection. *International Journal of Intelligent Systems*, 37.

McMahan, B.; and Ramage, D. 2017. Federated Learning: Collaborative Machine Learning without Centralized Training Data.

Nishio, T.; and Yonetani, R. 2019. Client Selection for Federated Learning with Heterogeneous Resources in Mobile Edge. In *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*. IEEE.

Pandey, S. R.; Nguyen, L. D.; and Popovski, P. 2022. A Contribution-based Device Selection Scheme in Federated Learning. *IEEE Communications Letters*, 1–1.

Sasindran, Z.; Yelchuri, H.; and Prabhakar, T. V. 2023. Ed-Fed: A generic federated learning framework with resource-aware client selection for edge devices. In *2023 International Joint Conference on Neural Networks (IJCNN)*. IEEE.

Sattler, F.; Wiedemann, S.; Müller, K.-R.; and Samek, W. 2019. Robust and Communication-Efficient Federated Learning from Non-IID Data. arXiv:1903.02891.

Tahir, M.; and Ali, M. I. 2022. On the Performance of Federated Learning Algorithms for IoT. *IoT*, 3: 273–284.

Wang, S.; Tuor, T.; Salonidis, T.; Leung, K. K.; Makaya, C.; He, T.; and Chan, K. 2019a. Adaptive Federated Learning in Resource Constrained Edge Computing Systems. arXiv:1804.05271.

Wang, X.; Han, Y.; Wang, C.; Zhao, Q.; Chen, X.; and Chen, M. 2019b. In-Edge AI: Intelligentizing Mobile Edge Computing, Caching and Communication by Federated Learning. *IEEE Network*, 33: 156–165.

Zawad, S.; Ali, A.; Chen, P.-Y.; Anwar, A.; Zhou, Y.; Baracaldo, N.; Tian, Y.; and Yan, F. 2021. Curse or Redemption? How Data Heterogeneity Affects the Robustness of Federated Learning. *arXiv:2102.00655*, v1.

Zhang, S. Q.; Lin, J.; and Zhang, Q. 2022. A Multi-agent Reinforcement Learning Approach for Efficient Client Selection in Federated Learning. arXiv:2201.02932.

Zhang, X.; Hong, M.; Dhople, S.; Yin, W.; and Liu, Y. 2020. FedPD: A Federated Learning Framework with Optimal Rates and Adaptivity to Non-IID Data. *arXiv:2005.11418*, v3.