Adaptive Homogeneity-Based Client Selection Policy for Federated Learning

Yiming Xie*, Pinrui Yu*, Geng Yuan†, Xue Lin*, Ningfang Mi*,
*Department of Electrical and Computer Engineering, Northeastern University, Boston, USA
†School of Computing, University of Georgia, Athens, USA

2024 International Symposium on Networks, Computers and Communications (ISNCC) | 979-8-3503-6491-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISNCC62547.2024.10759031

***Abstract*—Federated learning (FL) is a distributed paradigm that enables multiple clients or edge devices to collaboratively train a model without sharing their local data. The FL system has to tackle a significant challenge due to the non-IID nature of client data. Traditional methods of client selection usually suffer from the problem of variable test accuracies as well as slow convergence because of their incapability in effectively handle data heterogeneity across clients. In this paper, we propose an Adaptive Homogeneity-Based Client Selection Policy (ASTraFL) to address this challenge. ASTraFL dynamically selects clients whose data distributions optimally complement the current state of the global model, focusing on increased homogeneity of the selected client data in each training round. Our experiments demonstrate that ASTraFL can accelerate convergence speed and ensure the learning process's robustness.**

***Index Terms*—federated learning, client selection, homogeneity, robustness, non-IID**

I. Introduction

In recent years, federated learning (FL) has emerged as a prominent decentralized machine learning paradigm, enabling multiple parties to collaboratively train models without sharing their local data [1]–[3]. This approach not only preserves data privacy and security but also utilizes distributed computational resources efficiently. Recent advancements in FL research have focused on enhancing model accuracy, communication efficiency, and data security. Techniques such as model aggregation algorithms, optimized communication protocols, and differential privacy have propelled federated learning to the forefront of applicable solutions in areas requiring stringent data confidentiality and efficiency, including healthcare, finance [4] and telecommunications [5], heralding a new era of privacy-preserving, decentralized artificial intelligence.

One of the fundamental challenges in FL is the non-Independent and Identically Distributed (non-IID) nature of data across different clients [6], [7]. In practical scenarios, the data collected by each client can vary significantly in terms of distribution, quantity, and quality, leading to skewed datasets. This heterogeneity poses significant challenges in model training, as algorithms designed for IID data often perform poorly on non-IID data [8], resulting in biased models that favor certain data distributions over others. Addressing the non-IID data challenge in FL has spurred the development of innovative strategies, as highlighted in recent research. Client selection mechanisms, such as those proposed by [9]–[11], aim to enhance model training by optimizing client participation based on data distribution and computational capabilities. Techniques to combat data skewness include data augmentation methods [12], which diversify training datasets across clients using cropping, rotation, and flipping to improve model robustness. Specialized model aggregation methods, like FedShare [13] and SkewScout [14], focus on adjusting model updates to better accommodate the variance in data distributions. Additionally, efforts [15], [16] to ensure fairness and reduce bias in federated models emphasize equitable resource allocation and model performance across diverse client data. These strategies collectively address the complexities of training accurate and fair models in the inherently heterogeneous environments characteristic of federated learning.

Traditional client selection methods in FL have primarily relied on either random selection [2], [17] or strategies that target specific metrics such as data volume or loss magnitudes [18], [19]. Some studies have explored the importance of sampling to improve the learning process by prioritizing updates from more influential clients. For instance, [20] utilizes metrics like larger gradient norms and defines an upper bound for selected samples, aiming to accelerate convergence by optimally balancing communication resources and data selection. However, these methods often neglect the underlying data distribution patterns across clients, leading to inconsistent test accuracies and inefficient learning processes.

To address this issue, this paper introduces an Adaptive Homogeneity-Based Client Selection Policy (ASTraFL), which significantly diverges from conventional strategies by dynamically learning and adapting to the selection pattern that best aligns with the global model's needs. Our new method considers not only the quantity of data at each client but also the representativeness of the overall dataset, thus addressing the challenge of non-IID data more effectively.

Our contributions are summarized as follows:

- We propose a novel client selection approach that adapts to the inherent data distribution characteristics of each client, promoting faster convergence and more robust model training.
- Our approach, ASTraFL, builds upon traditional selection patterns and further incorporates real-time data homogeneity analysis, enabling a more informed and strategic selection process.
- We conduct extensive experiments to show that our approach outperforms existing methods, achieving higher accuracy and more stable convergence across a variety of non-IID settings.

979-8-3503-6491-0/24/31.00 © 2024IEEE

- Our study provides insights into the effectiveness of adaptive selection strategies in federated learning, offering a solution for more efficient selection methods in heterogeneous data environments.

In the remainder of this paper, we introduce the background and motivation in Sec. II. We present the detailed selection policy and the algorithms of ASTraFL in Sec. III. Sec. IV evaluates ASTraFL across different non-IID data settings and Sec. V presents the conclusion and future work.

II. Motivation

Federated learning inherently presents a unique set of challenges when it comes to selecting clients to participate in model training, particularly in scenarios with non-IID data distributions. Traditional methods such as random selection or focusing on high-loss clients often yield inconsistent results. Our study, conducted with a simple CNN model using the CIFAR-10 dataset across 12 clients under various conditions, provides insights into the selection behaviors of these methods.

Specifically, we use the CIFAR-10 dataset distributed under different Dirichlet distributions (i.e., non-IID datasets) and a uniform distribution (i.e., IID datasets) to evaluate different client selection strategies. For example, Figure 1 depicts the label distributions of the datasets with (a) a uniform distribution (denoted as $\mathcal{U}$), (b) a Dirichlet distribution with $\beta = 0.5$ [1]; and (c) a mixed Dirichlet distribution with different values of β, (denoted as β^*). In the mixed case, we set one third of the clients with β equal to 0.1, 0.4, and 0.9. We can observe that different data distributions can influence the availability of labels per client, thus affecting client selection strategies.

We then conduct the following three traditional client selection strategies on the datasets with different data distributions: (1) *Random* - Clients are chosen at random for each round of the learning process; (2) *High Loss* - Clients with the highest training loss are prioritized; and (3) *Cluster* - Clients are grouped based on the similarity of their data distributions and then representatives from each cluster are selected. The selection ratio is 50%. These traditional strategies will also be used as the baselines for the comparisons in our evaluation. Table I shows the testing accuracies (including min, max, and average) and the corresponding standard deviations under various client selection policies. Figure 2 illustrates the convergence of testing accuracies across different communication rounds and Figure 3 shows the selected client distributions when we have the mixed Dirichlet distribution β^*.

As shown in Table I and Figure 2, the *High Loss* method performs better than *Random* under a uniform distribution, as the equal distribution of labels allows this method to effectively enhance learning in each training round. However, under non-uniform distributions such as $\beta = 0.5$ and β^*, where label distribution is skewed, the efficacy of *High Loss* diminishes because prioritizing high-loss clients under skewed distributions may not optimally contribute to model learning due to the uneven representation of labels. In contrast, the *Cluster* method outperforms *High Loss* and achieves similar results to *Ramdom* but with less variance, indicating better stability and robustness. Particularly in the mixed and challenging case of β^*, *Cluster* selects representative clients from each cluster to ensure a balanced contribution from various data distributions, see Figure 3c. These observations highlight the challenges associated with mixed Dirichlet distributions, which can exacerbate training difficulties and increase test accuracy variance. Our research thus focuses on this challenge case to identify optimal solution under various data distributions.

[1]The β value in a Dirichlet distribution controls the concentration of data across different categories. A β value (close to 1) typically indicates uniform distribution and more balanced data, while a β closer to 0 leads to extreme values, making the data more sparse and skewed.

TABLE I: Performance metrics across different distributions and policies.

Distribution	Policy	Min Acc	Max Acc	Avg Acc	Std ($\times 10^{-3}$)
$\mathcal{U}$	Random	63.25	64.63	64.12	3.02
	High Loss	64.66	65.83	65.29 ↑	3.69 ↑
	Cluster	64.65	65.40	65.12 ↓	2.03 ↓
$\beta = 0.5$	Random	57.72	62.26	60.19	11.67
	High Loss	55.83	60.54	58.36 ↓	13.23↑
	Cluster	58.93	61.27	60.65 ↑	7.59 ↓
β^*	Random	51.54	61.15	57.91	22.19
	High Loss	50.35	58.61	53.58 ↓	22.47 ↑
	Cluster	53.46	61.36	58.51 ↑	24.97 ↑

III. Design of ASTraFL

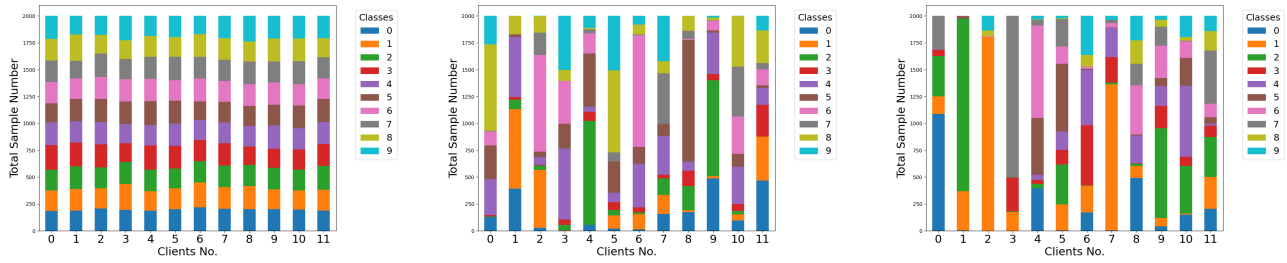
In this section, we present the design of our new client selection policy, named ASTraFL. The goal of ASTraFL is to effectively address the challenge of non-IID datasets, particularly with mixed Dirichlet distributions across different clients.

TABLE II: Summary of Notations

Symbol	Description
K_j	Cluster j, where j is the cluster number.
H_{K_j}	Average homogeneity of cluster K_j.
c_i	Client i.
D_i	Client c_i's dataset.
h_{c_i}	Homogeneity measure for client c_i by $\mathcal{H}(D_i)$.
δ	Homogeneity threshold.
F_{K_j}	Ratio of the total homogeneity within the cluster K_j to the total homogeneity across all clients.
N	Total number of available clients.
N_S	Total number of clients selected for training.
P	Selection percentage.
$N_S^{K_j}$	Number of clients selected from cluster K_j.

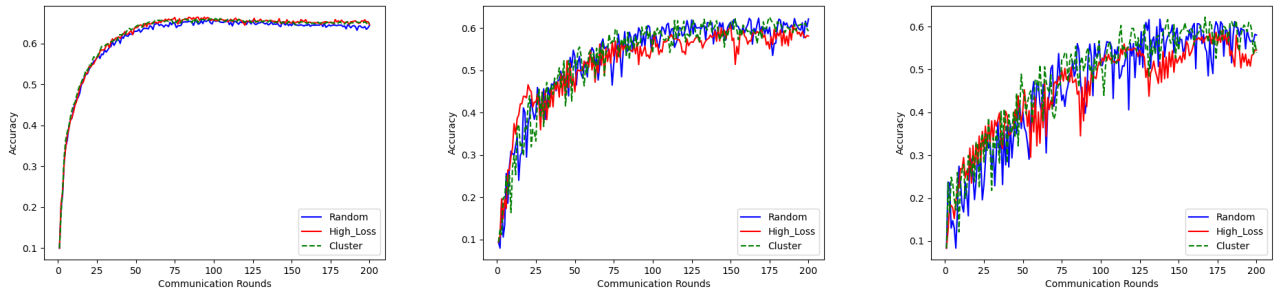
A. Problem Formulation

Consider a federated learning system consisting of a central server and a set of N clients, each possessing its own local dataset D_i. The objective is to construct a global model M with parameters θ, using locally computed updates to minimize a global loss function $\mathcal{L}(\theta)$. The challenge lies in selecting client updates that can improve the overall performance of the model, especially in the presence of data heterogeneity. Table II summarizes the key notations used in this paper.



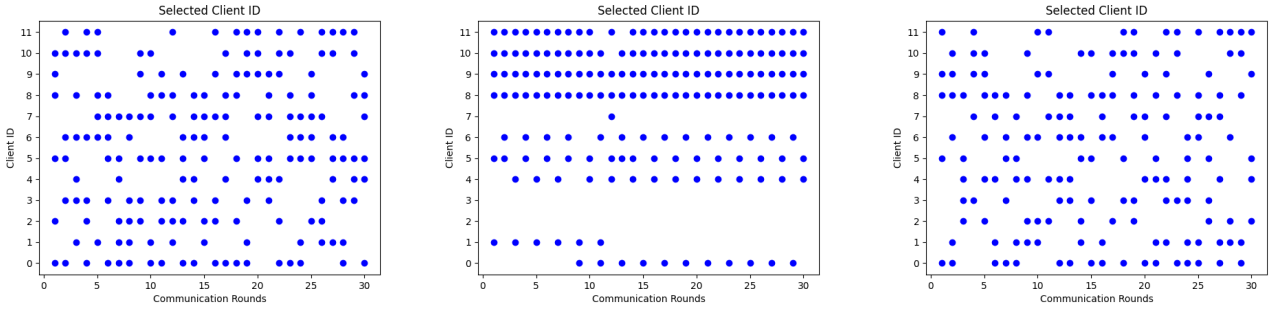
(a) $\mathcal{U}$ (b) $\beta = 0.5$ (c) β^*

Fig. 1: Label distributions under different Dirichlet distributions.



(a) $\mathcal{U}$ (b) $\beta = 0.5$ (c) β^*

Fig. 2: Test accuracy under different data distributions.



(a) *Random* (b) *High Loss* (c) *Cluster*

Fig. 3: Patterns of different selection methods under β^*.

B. Client Selection Policy - ASTraFL

In ASTraFL, each client is assessed based on three key metrics: Negative Log-Likelihood (NLL) homogeneity, loss rate, and content diversity of the dataset. **NLL Homogeneity**: This metric indicates the degree to which the local model aligns with the data from a client. **Loss Rate**: This metric reveals the rate at which the local model's loss is decreasing, providing insights into the client's learning efficiency. **Content Diversity**: This metric [21] evaluates the diversity within a client's dataset. A higher content diversity indicates that the model learns more generalizable features applicable across various domains. These three metrics are computed for each client in every communication round.

C. Loss Function

The loss function for each client aims to capture both the model's fit to the client's data and the homogeneity of this data in relation to the global dataset. It is expressed as follows:

$$\mathcal{L}(\theta) = \text{NLL}(\theta; D_i) + \lambda \cdot \mathcal{H}(D_i), \tag{1}$$

where $\text{NLL}(\theta; D_i)$ represents the Negative Log-Likelihood of the model parameters θ given the client's dataset D_i, reflecting how well the model fits. $\mathcal{H}(D_i)$ quantifies the homogeneity of the client's dataset in relation to the global data distribution, and λ is a regularization parameter used to balance these two terms.

D. NLL Computation

The NLL for a client c_i is defined to evaluate the model's fit to the client's specific data distribution, given by:

$$\text{NLL}_{c_i}(\theta) = -\sum_{x \in D_i} \log p_\theta(y|x), \tag{2}$$

where D_i represents the dataset of client c_i, x are the data points, y are the corresponding labels, and p_θ is the model's probability output parameterized by θ.

E. Adaptive Selection Based on Homogeneity

Within a cluster K_j, we define the average homogeneity as:

$$H_{K_j} = \frac{1}{|K_j|} \sum_{c_i \in K_j} h_{c_i}, \tag{3}$$

where h_{c_i} measures the homogeneity of client c_i's data characteristics. We also introduce a metric F_{K_j} to evaluate the significance of each cluster K_j to the network's overall homogeneity. It is defined as the ratio of the total homogeneity within the cluster to the total homogeneity across all clients in each communication round:

$$F_{K_j} = \frac{\sum_{c_i \in K_j} h_{c_i}}{\sum_{c_i \in N} h_{c_i}}. \tag{4}$$

F_{K_j} plays as a critical role in determining the selection priority of clusters for update aggregation into the global model. Specifically, clusters with higher F_{K_j} values, indicating a larger share of the network's total homogeneity, are prioritized in the selection process, e.g., more clients can be selected from such clusters.

Additionally, the selection criterion $S(K_j)$ for each cluster K_j is updated based on its individual homogeneity metric H_{K_j} as follows:

$$S(K_j) = \begin{cases} \text{high loss selection} & \text{if } H_{K_j} > \delta, \\ \text{content diversity selection} & \text{otherwise.} \end{cases} \tag{5}$$

Here, δ is defined as the homogeneity threshold. This strategy ensures that updates from clusters that exceed δ and also represent a significant portion of the network's homogeneity are prioritized by selecting clients with higher loss rates. Conversely, clients with higher content diversity will be chosen from the clusters with less homogeneity, aiming to improve training accuracy.

F. Algorithm Overview

The overview of ASTraFL is presented in Algorithm 1. This algorithm iterates through communication rounds, where each round involves computing three key metrics (i.e., NLL homogeneity, loss rate, and content diversity) for each client, applying adaptive selection schemes to each cluster, selecting client updates, and aggregating them into the global model. This process ensures efficient and effective training of the global model, utilizing the inherent heterogeneity of data across clients to enhance overall performance.

Algorithm 1 ASTraFL- Adaptive Homogeneity-Based Client Selection Policy Overview

Input: Initialize federated learning system with clients N with datasets $D = \{D_1, D_2, \ldots, D_N\}$
Output: Selected clients $S = \{S_1, S_2, \ldots, S_{N_S}\}$
foreach *communication round* **do**
  foreach *client* $c_i \in N$ **do**
    Compute h_{c_i}, loss rate, and content diversity
  end
  Group clients into clusters $\{K_1, K_2, \ldots, K_j\}$ based on cosine similarity
  foreach *cluster* K_j **do**
    Compute average homogeneity metric H_{K_j}
    Compute selected numbers $N_S^{K_j}$ as described in Algorithm 2
    if $H_{K_j} > \delta$ **then**
      Apply high loss selection policy to K_j
    else
      Apply content diversity selection policy to K_j
    end
  end
end

Algorithm 2 Compute Cluster Client Selection Numbers

Input: Set of clusters $\{K_1, K_2, \ldots, K_j\}$, homogeneity measures $\{h_{c_i}\}$ for all clients c_i, Selection Percentage P, N
Output: Number of clients selected from each cluster $N_S^{K_j}$
foreach *cluster* K_j **do**
  Calculate average homogeneity H_{K_j} across all clients
  Compute fraction F_{K_j} for cluster K_j
  Calculate number of clients to select:

$$N_S^{K_j} = \lceil F_{K_j} \times N \times P \rceil$$

end

IV. Evaluation

A. Experimental Setup

Our evaluation framework is implemented on IBM's Federated Learning (IBMFL) [22] platform, leveraging a powerful computational environment equipped with four NVIDIA GeForce GTX 1080 Ti GPUs. To benchmark the performance of federated learning models, particularly their effectiveness against non-IID data distribution, we use CIFAR-10 as the dataset to generate various non-IID data distribution scenarios.

Two neural networks are chosen in the evaluation: (1) **Conv-3** model that consists of three convolutional layers with 32, 64, and 128 filters, each of size 3×3 followed by a 2×2 max-pooling layer, and concludes with a flattening layer, two dense layers with 256 and 128 neurons, respectively, and an output layer with softmax activation to classify the classes; and (2) **ResNet-20** model that is based on the ResNet architecture, known for its effectiveness in image classification tasks. Specifically, ResNet-20 variant comprises residual blocks with 20 layers.

B. Workload Design

To thoroughly test the models' capabilities and simulate real-world scenarios, we design four distinct workloads involving 12 clients each, where every client exclusively draws 2000 samples from the CIFAR-10 dataset:

- **W1:** This workload involves clients with data distributions represented by $\beta^* = [0.1, 0.4, 0.5, 0.8]$, distributed

evenly with a ratio of 3:3:3:3. It serves as a baseline to assess model performance under balanced data distributions.

- **W2:** In this workload, the data distributions remain the same as in **W1** ($\beta^* = [0.1, 0.4, 0.5, 0.8]$), but with a skewed ratio of 1:5:5:1. This workload tests the robustness of models under scenarios where certain classes dominate the dataset.
- **W3:** The data distributions are adjusted to $\beta^* = [0.5, 0.6, 0.7, 0.8]$, maintaining an even ratio of 3:3:3:3. The focus is on introducing higher diversity within similar distributions.
- **W4:** This workload has the data distribution with lower $\beta^* = [0.1, 0.2, 0.3, 0.4]$ with a balanced ratio of 3:3:3:3. It aims to explore the impact of lesser variability among lower β values on model performance.

These workloads are designed to cover a wide range of potential uncertainties in data distribution encountered by local clients. By incorporating variations in skewness, diversity, and variability, these workloads enable a comprehensive evaluation of the adaptability and performance of the models in the diversified federated environment.

C. Results

Tables III and IV provide a summary of the test accuracy results obtained from different workload scenarios for Conv-3 and ResNet-20, respectively. Figure 4 shows the runtime test accuracies under four selection policies for Conv-3. Notably, in scenario **W1**, our method obtains the highest average accuracy compared to other methods. This outcome can be attributed to our strategic approach, which prioritizes the selection of more homogeneous clients (8-11), moderately selects clients (4-7), and minimizes selection from highly biased clients (0-3). The corresponding client selections of ASTraFL can be observed in Figure 5a.

When a workload, such as **W2**, contains certain classes dominating the dataset, our selection pattern prioritizes clients with data distribution of $\beta = 0.4$ and $\beta = 0.5$, see Figure 5b. This methodological preference assures that the data from the major distributions are embraced in training our model. By focusing on these dominant distributions, ASTraFL improves the model's ability to effectively capture and learn from the most prevalent classes in the dataset.

In **W3**, where an even ratio with high β indicates less bias in the data distributions, we observe a balanced selection across all β values, see Figure 5c. This balanced selection ensures that contributions are made across the entire range of data distributions, enhancing the overall robustness and accuracy of the model. As shown in Table III, higher accuracy and lower variance are obtained under all policies compared to **W1** and **W2** and ASTraFL again achieves the best test accuracy under this workload.

W4, on the other hand, has the most biased data distributions. In Table III, we observe the lowest accuracies and the highest standard deviations of accuracy under all three baselines (i.e., Random, High Loss and Cluster). However, our approach remains competitive despite this fact, as illustrated in Figure 4d. Meanwhile, our selection method optimizes toward content diversity, effectively seeking informative clients and further improving learning from a broader representation of the class, even in scenarios with biased data distributions.

Finally, in Figure 6 we plot the heatmap of optimal cluster numbers by ASTraFL under **W1**, where the left y-axis shows communication rounds and the heat value indicates different numbers of clusters. For example, we have five clusters at round 2, but only two clusters at round 48. Compared to traditional clustering approaches, Figure 6 illustrates how the dynamic adaptation of ASTraFL enables continuous fine-tuning of the model. This adaptability ensures its ability to adapt to evolving data insights, thus confirming the effectiveness of our approach in addressing increasingly diverse and challenging environments in federated learning.

TABLE III: Performance metrics across different distributions and policies for Conv-3.

Workload	Policy	Min Acc	Max Acc	Avg Acc	Std ($\times 10^{-3}$)
$\mathbf{W}_1$	*Random*	50.97	58.63	54.59	20.66
	High Loss	49.38	54.57	52.77↓	16.19↓
	Cluster	47.68	57.09	52.87↓	21.52↑
	ASTraFL	52.85	59.67	57.32↑	18.09↓
$\mathbf{W}_2$	*Random*	52.68	57.85	55.54	14.32
	High Loss	53.15	56.79	54.65↓	10.10↓
	Cluster	52.89	58.74	56.14↑	17.4 ↑
	ASTraFL	54.80	59.70	57.99↑	12.68↓
$\mathbf{W}_3$	*Random*	56.86	60.69	59.32	9.79
	High Loss	57.64	59.61	58.58↓	5.49↓
	Cluster	59.35	61.69	60.53↑	6.80↓
	ASTraFL	60.99	63.84	62.61↑	8.67↓
$\mathbf{W}_4$	*Random*	41.62	55.07	50.08	33.55
	High Loss	41.23	53.00	49.03↓	26.96↓
	Cluster	43.87	55.33	51.41↑	27.77↓
	ASTraFL	46.47	58.51	54.20↑	33.12↓

TABLE IV: Performance metrics across different distributions and policies for ResNet-20.

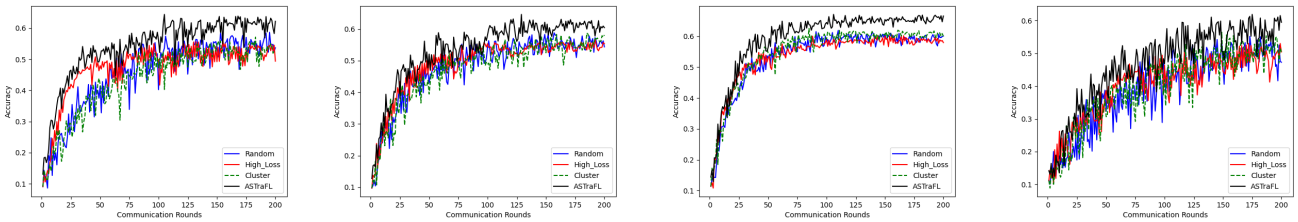
Workload	Policy	Min Acc	Max Acc	Avg Acc	Std ($\times 10^{-3}$)
$\mathbf{W}_1$	*Random*	25.38	50.47	40.84	70.45
	High Loss	30.94	53.73	43.82↑	60.03↓
	Cluster	19.74	50.44	39.12↓	67.11↑
	ASTraFL	41.54	59.16	48.07↑	47.02↓

V. Conclusion

This paper introduces ASTraFL, a novel client selection approach designed to address the challenges associated with both IID and non-IID data distributions. By intelligently selecting clients based on data homogeneity, ASTraFL is able to enhance coherence and relevance in updates aggregation during model training and thus achieves higher test accuracies than traditional strategies under various data distributions.In the future, we will explore scaling up the number of clients to assess the robustness and efficiency of our policy across an even broader and more diverse data landscape.

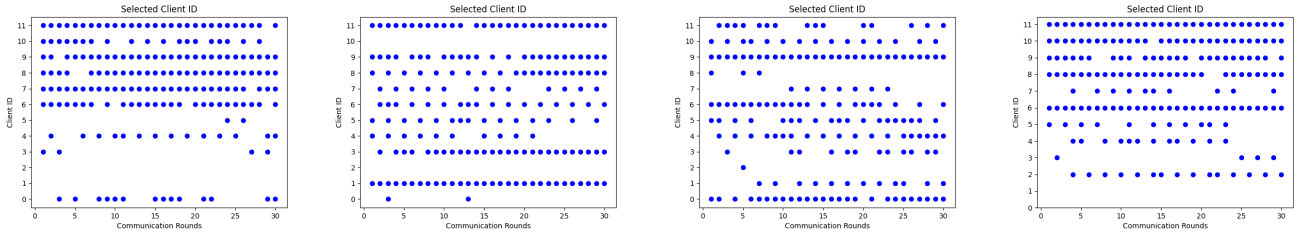
VI. Acknowledgments

This work was partially supported by National Science Foundation Award CNS-2008072.



(a) $\mathbf{W}_1$ (b) $\mathbf{W}_2$ (c) $\mathbf{W}_3$ (d) $\mathbf{W}_4$

Fig. 4: Test accuracies under four different workloads and policies for Conv-3.



(a) $\mathbf{W}_1$ (b) $\mathbf{W}_2$ (c) $\mathbf{W}_3$ (d) $\mathbf{W}_4$

Fig. 5: Client selection patterns by ASTraFL under four different workloads for Conv-3.

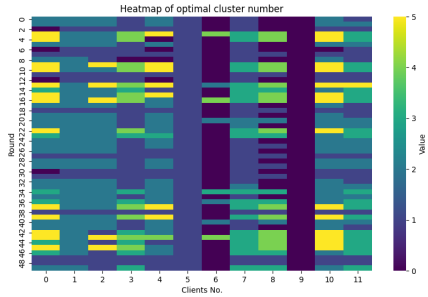


Fig. 6: Heatmap of optimal cluster numbers by ASTraFL under $\mathbf{W_1}$ for Conv-3.

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] J. Konečnỳ, B. McMahan, and D. Ramage, "Federated optimization: Distributed optimization beyond the datacenter," *arXiv preprint arXiv:1511.03575*, 2015.

[3] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE signal processing magazine*, vol. 37, no. 3, pp. 50–60, 2020.

[4] T. Liu, Z. Wang, H. He, W. Shi, L. Lin, R. An, and C. Li, "Efficient and secure federated learning for financial applications," *Applied Sciences*, vol. 13, no. 10, p. 5877, 2023.

[5] N. H. Tran, W. Bao, A. Zomaya, M. N. H. Nguyen, and C. S. Hong, "Federated learning over wireless networks: Optimization model design and analysis," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, 2019, pp. 1387–1395.

[6] X. Ma, J. Zhu, Z. Lin, S. Chen, and Y. Qin, "A state-of-the-art survey on solving non-iid data in federated learning," *Future Generation Computer Systems*, vol. 135, pp. 244–258, 2022.

[7] H. Zhu, J. Xu, S. Liu, and Y. Jin, "Federated learning on non-iid data: A survey," *Neurocomputing*, vol. 465, pp. 371–390, 2021.

[8] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, "Scaffold: Stochastic controlled averaging for federated learning," in *International conference on machine learning*. PMLR, 2020, pp. 5132–5143.

[9] 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.

[10] H. Wang, Z. Kaplan, D. Niu, and B. Li, "Optimizing federated learning on non-iid data with reinforcement learning," in *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications*, 2020, pp. 1698–1707.

[11] D. Y. Zhang, Z. Kou, and D. Wang, "Fedsens: A federated learning approach for smart health sensing with class imbalance in resource constrained edge computing," in *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications*, 2021, pp. 1–10.

[12] Y. Yan and L. Zhu, "A simple data augmentation for feature distribution skewed federated learning," *arXiv preprint arXiv:2306.09363*, 2023.

[13] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated learning with non-iid data," *arXiv preprint arXiv:1806.00582*, 2018.

[14] K. Hsieh, A. Phanishayee, O. Mutlu, and P. Gibbons, "The non-iid data quagmire of decentralized machine learning," in *International Conference on Machine Learning*. PMLR, 2020, pp. 4387–4398.

[15] G. Wang, A. Payani, M. Lee, and R. Kompella, "Mitigating group bias in federated learning: Beyond local fairness," *arXiv preprint arXiv:2305.09931*, 2023.

[16] D. Y. Zhang, Z. Kou, and D. Wang, "Fairfl: A fair federated learning approach to reducing demographic bias in privacy-sensitive classification models," in *2020 IEEE International Conference on Big Data (Big Data)*, 2020, pp. 1051–1060.

[17] C. Li, X. Zeng, M. Zhang, and Z. Cao, "Pyramidfl: A fine-grained client selection framework for efficient federated learning," in *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*, 2022, pp. 158–171.

[18] A. H. Jiang, D. L.-K. Wong, G. Zhou, D. G. Andersen, J. Dean, G. R. Ganger, G. Joshi, M. Kaminksy, M. Kozuch, Z. C. Lipton *et al.*, "Accelerating deep learning by focusing on the biggest losers," *arXiv preprint arXiv:1910.00762*, 2019.

[19] H.-S. Chang, E. Learned-Miller, and A. McCallum, "Active bias: Training more accurate neural networks by emphasizing high variance samples," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[20] Y. He, J. Ren, G. Yu, and J. Yuan, "Importance-aware data selection and resource allocation in federated edge learning system," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 13 593–13 605, 2020.

[21] A. Li, L. Zhang, J. Tan, Y. Qin, J. Wang, and X.-Y. Li, "Sample-level data selection for federated learning," in *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications*, 2021, pp. 1–10.

[22] H. Ludwig, N. Baracaldo, G. Thomas, Y. Zhou, A. Anwar, S. Rajamoni, Y. Ong, J. Radhakrishnan, A. Verma, M. Sinn *et al.*, "Ibm federated learning: an enterprise framework white paper v0. 1," *arXiv preprint arXiv:2007.10987*, 2020.