Heterogeneous FL via active-passive collaboration

Jianzhe Zhao[1], Wuganjing Song[2], Li Zhang[1], Xingwei Wang[3,4], Guibing Guo[1], and Zhelin Fan[1](✉)

[1] Software College, Northeastern University, Chuangxin Road 195, Shenyang, Liaoning 110169, China
{zhaojz, zhangl, guogb}@swc.neu.edu.cn, 20216976@stu.neu.edu.cn

[2] The Hong Kong University of Science and Technology
wsongan@connect.ust.hk

[3] College of Computer Science and Engineering, Northeastern University, Shenyang 110819, China

[4] State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819, China
wangxw@mail.neu.edu.cn

Abstract. Federated Learning (FL) is a privacy-preserving machine learning framework that often faces performance challenges due to data and system heterogeneity. Current client selection strategies improve heterogeneous FL performance but face two main challenges. Our preliminary experiments show that data heterogeneity arises not only globally but also within individual clients due to local data volume differences. However, typical server-side client selection strategies cannot fundamentally address accuracy declines from local data heterogeneity. Additionally, clients with high-quality data typically require longer training times, creating a trade-off between accuracy and training duration, and many strategies exclude these clients due to time delays, further reducing accuracy. To address these issues, we propose FCH-FL, enhancing heterogeneous FL through active-passive collaboration. Our method includes a client-side data balancing algorithm to reduce the impact of local data heterogeneity on model generalization and a server-side passive client selection strategy with a utility function that optimizes both time and accuracy, ensuring that high-quality clients are not excluded. Experiments on CIFAR-10 and EMNIST confirm the method's effectiveness: our approach improves global accuracy by up to 14.42% and 7.82% in Non-IID settings and achieves up to 1.78× and 1.45× gains in time efficiency compared to baselines.

Keywords: Federated learning · Data and system heterogeneity · Long-tail learning · Client selection.

1 Introduction

Federated Learning (FL) is a distributed machine learning paradigm designed to train models on local devices while protecting data privacy [10]. Due to its merits, it has widespread applications in fields like healthcare [15] and the IoT [12].

However, FL faces data and system heterogeneity challenges in practical applications. Data heterogeneity causes local data to deviate from the independent and identically distributed (IID) assumption, affecting the accuracy of the global model; system heterogeneity arises from differences in device hardware and network connectivity, leading to imbalances in computation and communication. Existing research optimizes the time-to-accuracy metric through client selection strategies to balance statistical and system utility while facing client selection challenges. It often overlooks two personalized scenarios: (1) Non-IID issues exist not only in variations between clients but also within individual clients' local data, such as long-tail local datasets. (2) Clients with high-quality data consume excessive time, leading to a trade-off between time and data quality as optimization objectives, potentially sacrificing statistical utility.

Drawing from experiments, we have three key observations. First, non-IID data significantly impact global accuracy, primarily due to the long-tailed distribution of local data and statistical bias from uneven label distribution. We performed experiments on EMNIST and found a positive correlation: the higher the IF(imbalance factor), the less pronounced the long tail, and the higher the global accuracy, as shown in Fig. 1(a-1). Then, adjusting skewness via a Dirichlet distribution shows that increased label distribution skewness results in a notable decline in global accuracy, as shown in Fig. 1(a-2). These two forms of data heterogeneity may interact, further reducing global accuracy. Second, with highly skewed non-IID data, Oort, an advanced client selection algorithm, achieves similar global accuracy to FedAvg, but its client selection strategy favors non-long-tailed clients, thus not effectively improving accuracy, as seen in Fig. 1(b). Finally, the trade-off between data quality (accuracy contribution) and training time (data volume) leads Oort to exclude high-accuracy-contributing clients in favor of time efficiency, as seen in Fig. 1(c).

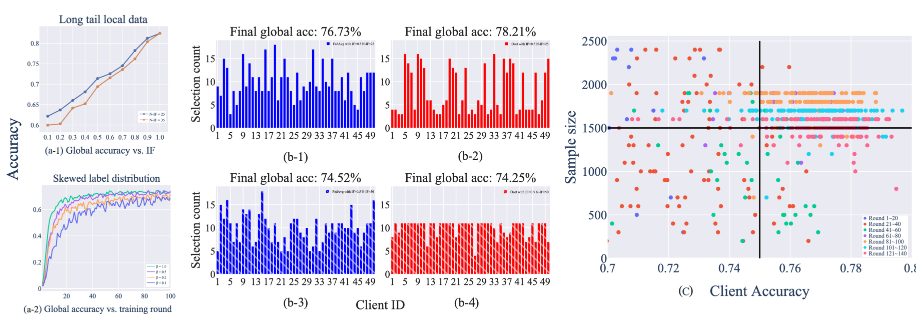

Fig. 1. Observations: (**a**) the effect of long tail local data and skewed label distribution on global accuracy (**b**) Oort's global accuracy performance and client selection in different degrees of client long tail (**c**) Oort excludes high-accuracy contribution clients from dilemma due to time penalty

Based on the observations, we believe introducing multi-level mechanisms to better balance data and system heterogeneity and optimize client selection criteria is key to enhancing time-to-accuracy. To this end, besides passive selection strategies, we implement a heterogeneous data balancing operation on the client side. This modifies the local model's Softmax cross-entropy using global data insights, reducing the negative impact of Non-IID data and improving the classification accuracy of the local model while preserving data privacy. We also refine client selection with a new utility function that comprehensively considers both types of heterogeneity, avoiding excluding high-quality data clients due to time efficiency and mitigating conflicts between utility optimization and accuracy. In summary, this study makes the following contributions.

- We propose FCH-FL, achieving fast convergence heterogeneous FL via active-passive collaboration. We focus on data and system heterogeneity problems to improve time-accuracy performance, especially in personalized scenarios.
- We introduce a heterogeneous data balancing algorithm, which uses prior knowledge of global data labels to adjust the local cross-entropy loss function, enhancing the generalization ability and improving the accuracy of long-tail local data while ensuring data privacy.
- We further improve the client selection strategy by defining a new client utility function that balances the potentially conflicting objectives of data and system heterogeneity. This enhancement allows us to optimize the trade-off between statistical and system utility in the FL system.
- We conduct comprehensive experiments on the CIFAR-10 and EMNIST datasets, considering data and system heterogeneity settings. We compare the performance of our algorithm against baselines, examining its efficiency, convergence, and accuracy.

2 Background

2.1 Problem formation

This study considers a heterogeneous FL system with K clients, denoted as $K = \{K_i \mid i = 1, 2, \ldots, n\}$, and a central server S. Each client maintains its own local dataset $D = \{D_i \mid i = 1, 2, \ldots, n\}$. The samples in the local dataset D_i can be represented as $\{x_j, y_j\}$, where x_j represents the feature vector of a data sample, and y_j represents the label of the data sample. In the FL process, after receiving the global model M, each client K_i trains the model using its local dataset D_i, with the objective function defined as follows: $f_i\left(w\right) = \frac{1}{|D_i|} \sum_{j=1}^{|D_i|} f\left(w, x_j, y_j\right)$. Besides, we define t_i as the time consumption measured from the moment the server S receives the returned models. We aim to design an algorithm A that minimizes the time required to achieve the global loss function $F(w)$ by leveraging local client operations and server-side selection of appropriate client subsets. This can be mathematically represented as Equ. 1.

$$\min\left(epoch\right) \qquad subject\ to \qquad \arg\min F\left(w\right)=\sum_{i=1}^{K}p_i f_i\left(w\right) \tag{1}$$

where $f_i\left(w\right)$ represents the loss value of k_i, $F\left(w\right)$ represents the global loss of the aggregated client models, and p_i represents the weight of its local model.

2.2 Heterogeneous federated learning

Data heterogeneity Non-IID data can take various forms. In this work, we focus on skewed label distribution among clients and the long-tail phenomenon of local data. We define data heterogeneity as the Non-IID nature among data D and particularly in D_i, which is identified as the primary factor for the decline in the utility of the global model in FL [8].

Definition 1 (Skewed Label Distribution). *For FL datasets D, let X denote the feature space $X=\{x_1,x_2,\ldots,x_n\}$, and Y represent the label space $Y=\{y_1,y_2,\ldots,y_n\}$. For two clients K_i and K_j, the distribution of the same label y_k in their respective local datasets D_i and D_j exhibits significant differences, as shown in Equ. 2*

$$p(y_{ki})\neq p(y_{kj}) \tag{2}$$

Definition 2 (Long-tail Local Data). *For the local dataset D_i of client K_i, the distribution of labels r and t on D_i exhibits significant differences, as shown in Equ. 3. In extreme cases, some samples dominate the local sample space while others are relatively scarce, which is known as the long-tail issue of local data [17].*

$$p\left(y_{ri}\right)\neq p\left(y_{ti}\right) \tag{3}$$

The learning objective of FL clients (Equ. 1) is to minimize the loss function $f\left(w\right)$. The most commonly used loss function is the Softmax CE, which measures the discrepancy between predicted results and true labels, as shown in Equ. 4.

$$f(w)=\ell(y,f(x))=\log\left[\sum_{y'\in[L]}e^{f_{y'}(x)}\right]-f_y(x) \tag{4}$$

System heterogeneity System heterogeneity is the main cause of heterogeneity in synchronous FL. In this context, the evaluation metric of interest is the minimum total wall-clock time. Besides, we assume the server has abundant computational resources compared to participants, thus ignoring the delay from global model aggregation. In this study, the time cost for participant i in t-th round of iterations depends on three main components: parameter broadcast time $\tau_i^D\left(t\right)$, local update time $\tau_i^L\left(t\right)$, and parameter upload time $\tau_i^U\left(t\right)$. Thus, the total time consumed by client i in t-th round can be expressed as Equ. 5.

$$\tau_i(t)=\min\{\tau_i^D(t)+\tau_i^L(t)+\tau_k^U(t),\tau_{max}\} \tag{5}$$

where t_{max} represents the maximum interval for communication in each round, which is used to avoid endless waiting caused by potential stragglers. In each communication round, the server waits until all selected participants have completed parameter upload before model aggregation. Thus, the time consumption of each round is determined by the slowest participant, given by Equ. 6.

$$\hat{\tau}(S(t)) = \max_{k \in S(t)} \tau_k(t) \tag{6}$$

Therefore, the problem of achieving the minimum time consumption for training the global model in FL can be formally defined as Equ. 7.

$$\min_{S(t), t \geq 1} \sum_{t=1}^{T} \hat{\tau}\left(S\left(t\right)\right) \tag{7}$$

3 Design of FCH-FL

3.1 Outline

In this work, we propose a novel algorithm to improve the trade-off between time and accuracy in heterogeneous federated systems. As shown in Fig. 2, the architecture FCH-FL consists of two key components: client-side active and server-side passive operations. On the client side, we introduce a client-side heterogeneous data balancing algorithm, which leverages global prior information to adjust the Softmax function of local models while preserving privacy. It resolves the issue of poor model generalization caused by long-tail data on the client side and mitigates the challenge of selecting appropriate clients in scenarios with a large number of Non-IID clients. On the server side, we enhance the existing client selection strategy, specifically addressing conflicting objectives of time and data quality. By refining the utility function, we avoid excluding high-quality data clients based solely on time efficiency. Additionally, we formulate the optimization problem as a multi-armed bandit (MAB) model to tackle the exploration-exploitation dilemma in client selection, effectively improving the efficiency of the selection algorithm.

3.2 Client-side heterogeneous data balancing algorithm

According to Def. 2, in long-tail local datasets, sample classification is challenging due to the majority class greatly outnumbering the minority class. Though the traditional solution introduces balanced cross-entropy, it still leads to unstable optimization and potential overfitting of the local model on minority samples. [11, 16]. To tackle this problem, Menon et al. [11] proposed a statistical approach that calibrates the Logit layer on top of the traditional Softmax CE. Based on their method, we introduce a client-side heterogeneous data balancing algorithm by incorporating prior estimates of the global dataset's label distribution into traditional cross-entropy loss. This approach enlarges the marginal distances of different class data in the sample space, compensating for cross-entropy

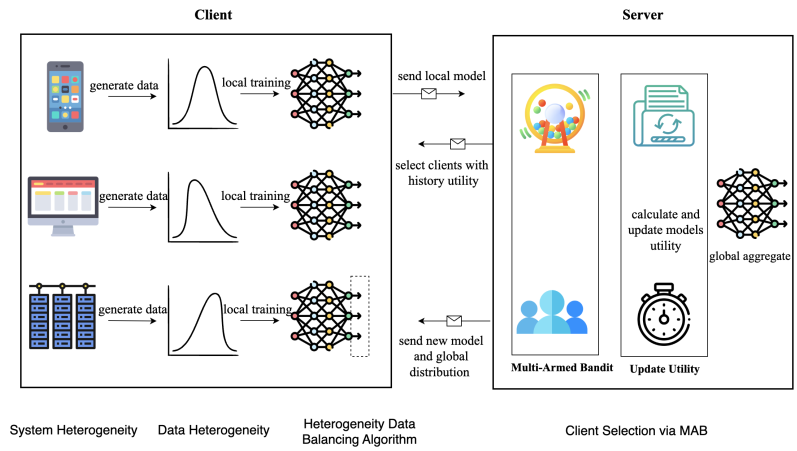

Fig. 2. FCH-FL architecture

limitations and allowing local models to have better generalization performance in heterogeneous data settings, ultimately enhancing the accuracy of a global model.

To implement the client-side heterogeneous data balancing algorithm, we propose the following local client model loss function in FL, as shown in Equ. 8.

$$l(y, f(x)) = -\log \frac{e^{f_y(x)+\tau\cdot\log\pi_y}}{\sum_{y'\in[Y]} e^{f_{y'}(x)+\tau\cdot\log\pi_{y'}}} \tag{8}$$

Here, π_y represents the global prior estimate of class y, $\pi_{y'}$ represents the global prior estimate of the negative class y', and τ is a hyperparameter that adjusts the scaling factor. For simplicity, we set $\tau = 1$ as the default value.

Menon et al. [11] proposed logit adjustment to modify the softmax CE to address class imbalance. According to their proof, Equ. 8 is Fisher consistent for minimizing balanced error, meaning it produces optimal predictions under class imbalance. In contrast, Equ. 4 lacks this consistency, so Equ. 8 is superior in solving the category imbalance problem. We generalize the logit adjustment to the FL setting. Comparing Equ. 4 and Equ. 8, the loss function in the client heterogeneity balancing algorithm introduces the prior estimate of class labels from global data obtained by statistically analyzing the global dataset distribution before clients' training. In our settings, a small portion of local data can be aggregated on the server to provide global prior information. Differential privacy can be applied during this process to enhance algorithm privacy [23]. Alg. 1 in the Appendix shows the client-side heterogeneous data balancing algorithm.

3.3 Server-side client selection strategy

To alleviate problems of heterogeneity, FCH-FL improves client selection. Firstly, a novel client utility function, denoted as U_t, is proposed to comprehensively quantify the statistical and system utility of client models towards the global model while incorporating necessary modifications for personalized scenarios. Next, the client selection problem is formulated as an MAB model, and the upper confidence bound (UCB) strategy is employed. The initial screening of all clients is based on their historical expected U_t and their number of participant rounds in global aggregation, aiming to balance the exploration-exploitation dilemma. Subsequently, the clients participating in the global aggregation are selected based on the magnitude of U_t in the current round of training, while a certain proportion of clients are randomly selected to enhance system robustness, thereby improving the time-to-accuracy of a global model.

Client utility function in heterogeneous FL. In FCH-FL, the server evaluates clients based on statistical and system utility, selecting those with the best overall performance for each iteration. Improving statistical utility reduces the number of iterations for convergence while enhancing system utility decreases average wait time. However, most clients excel in only one aspect. Existing methods like Oort tend to select clients with balanced but mediocre performance in both utilities, overlooking high-quality data clients(upper-right region of Fig. 1 (c)), which is crucial. To address limitations, we propose a new client selection function for heterogeneous FL. We prioritize statistical utility by selecting clients from the upper region and then adding well-performing clients from lower regions. A time regularization term is introduced to penalize clients with poor system utility, and we refine the selection criteria to retain clients in the upper-right region who significantly contribute to statistical utility, thereby enhancing the convergence performance of the global model. Furthermore, since the gradient vector represents the optimization direction and magnitude, we assume the closer a local model's gradient is to the global average gradient of the current round, the greater its contribution to global convergence [3]. Existing research has shown that the difference between global gradients of consecutive rounds in FL is minimal, allowing the previous round's global gradient to approximate the current round's one [4]. Thus, the difference between the previous round's global model gradient and the current round's local model gradient can approximate the local model's statistical utility, as shown in Equ. 9.

$$u_t(i) = \frac{1}{\| g_i - \overline{g_{t-1}} \|_2 + 1} \tag{9}$$

Here, $u_t(i)$ represents the statistical utility of the i-th client's model in the t-th round, g_i represents the gradient of the i-th client, and $\overline{g_{t-1}}$ represents the gradient of the global model in the previous $t-1$-th round. The use of the reciprocal ensures that a higher numerical value corresponds to a higher statistical utility of the local model.

The system utility primarily considers the training latency and penalizes client models with higher delays, as shown in Equ. 10.

$$\mathcal{R} = f\left(T, t, \alpha\right) = \begin{cases} \left(\dfrac{T}{t}\right)^{\alpha}, & if \quad T < t \\ 1, & if \quad T \geq t \end{cases} \tag{10}$$

Here, T represents the expected upper limit of global training time for the current round, which is dynamically adjusted by the server based on historical training and is typically set as the median of the time spent by all clients. α is the penalty factor parameter, which is set by the server before training and is usually set to 1. t represents the total time consumed by the current client in the current round. Only clients exceeding the server's time limit are penalized, while those finishing within the time aren't rewarded.

In conclusion, the local model utility $U_t(i)$ can be obtained using Equ. 11. We sort the client's statistical utility $u_t(i)$ and exempt the top $p\%$ high-quality clients from penalties. This ensures that high-quality clients with high time delays are not excluded from the training process.

$$U_t\left(i\right) = \begin{cases} u_t\left(i\right), & for \quad p\% \\ u_t\left(i\right) \cdot \mathcal{R}, & for \quad 1 - p\% \end{cases} \tag{11}$$

Client selection via MAB. In FL, thousands of clients participate in training, but the server can only gather a limited number at a time and select clients with better local model performance. In FCH-FL, we address the optimization problem by establishing an MAB model and balancing the exploration-exploitation trade-off through the UCB strategy. Alg. 2 in Appendix shows the details of Client selection via MAB algorithm.

4 Evaluation

4.1 Experimental setting

FCH-FL is designed for large-scale FL scenarios with heterogeneous clients varying in data and system characteristics. The key focus of FCH-FL is to maximize statistical utility while considering overall system utility, thereby improving FL performance. To evaluate the effectiveness of FCH-FL, we conducted a comprehensive set of experiments. We examined the improvement in accuracy achieved by FCH-FL and conducted ablation experiments to individually evaluate the impact of the client-side active balancing algorithm and server-side passive selection strategy on statistical utility enhancement. We also conducted extensive efficiency evaluations using the time-to-accuracy metric. Our experiments were performed with a CPU (AMD EPYC 7601, 16 cores, 64GB RAM) and a GPU (NVIDIA GTX 3090, 24GB VRAM).

Datasets and Models. CIFAR-10 [25] contains 60,000 RGB images (50,000 for training and 10,000 for testing), each 32x32 pixels, categorized into ten classes. The images feature a centered object with a simple background. **EMNIST** [2] contains 131,600 28x28 pixel images across 47 character categories from over 500 individuals.

Heterogeneity Setting. To model the data heterogeneity among clients, we segment the CIFAR-10 using the Dirichlet distribution [4] with the skew parameter set to 0.2. EMNIST dataset offers a Non-IID data partitioning that closely resembles real-world scenarios. We simulate the long-tail local data phenomenon using the standard metric IF ($IF = 0.5$) to measure the extent of the long tail in the data. Additionally, we introduce the parameter $\mu = 0.5$ to quantify the proportion of clients with long-tail data in the federated environment. We configure their processing latency to follow Zipf's distribution (with $\alpha = 1.2$) for the system heterogeneity setting.

Baselines. We compare the proposed FCH-FL algorithm with three baseline algorithms: FedAvg [10], FedProx [9], and Oort [7]. As the compared algorithms, we evaluate the performance of FCH-FL in terms of accuracy, convergence, and efficiency under the same hyperparameter settings.

- FedAvg. We perform FedAvg with clients randomly selected in evaluations. Each client trains a model locally and sends it to the server. The server randomly selects participants and averages their parameters to obtain the global model.
- FedProx. An advanced federated optimization algorithm with clients randomly selected in evaluations. It introduces an additional proximal term to enhance global model performance and achieves better results by constraining the Euclidean distance between local and global model parameters.
- Oort. An advanced client selection algorithm that improves performance by selecting clients based on their maximum loss value and faster training speed in each round.

Metrics. We use global model accuracy to assess the algorithm's performance. The loss values of the global model in each round are used to examine the convergence speed of the global model. The efficiency is evaluated using the time-to-accuracy metric, which quantifies the end-to-end duration of the testing process, including computational overhead and actual computation time.

4.2 Accuracy evaluation

We design a comprehensive evaluation of the FCH-FL's performance in IID and Non-IID scenarios, focusing on global model accuracy. We conduct ablation experiments to validate the advantages and adaptability of strategies for addressing the two data heterogeneity proposed in FCH-FL. Compared with FedAvg, Oort, and FedProx, we assess the abilities of the client-side heterogeneous balancing and server-side client selection strategy, which were sub-algorithms in FCH-FL, to mitigate accuracy reduction caused by data heterogeneity.

We initially evaluate the global accuracy performance of different algorithms under IID settings for both datasets. For CIFAR-10, we divide it into training and testing sets, with the training set comprising 80% of the total dataset. Subsequently, we allocate the training set to 50 clients following IID settings. In each round, 20% clients are selected, resulting in 200 training rounds. Similarly, we apply IID partitioning to EMNIST, evenly distributing its 131,600 samples among 200 clients. In each round, 10% of the clients are chosen to participate in global training, amounting to 100 training rounds. The global accuracy performance of different algorithms in IID environments is depicted in Fig. 3.

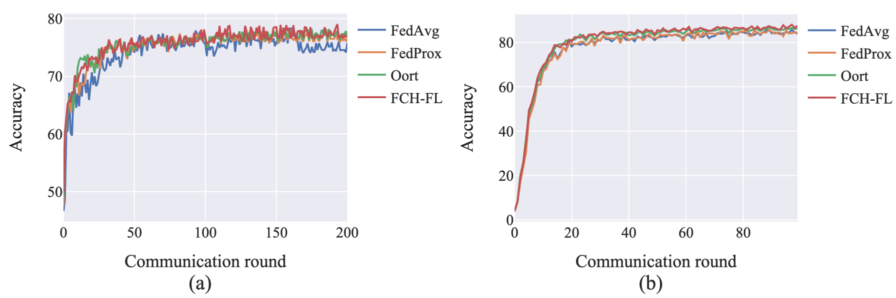

Fig. 3. Global model accuracy vs. training rounds for different FL frameworks on CIFAR-10 and EMNIST with IID partitioning: (**a**) CIFAR-10 and (**b**) EMNIST.

As shown in Fig. 3 (a), for CIFAR-10 under IID partitioning, FCH-FL shows some improvement in model accuracy but with a marginal difference compared to other algorithms. Specifically, FedAvg achieves the lowest accuracy at 74.13%, while FCH-FL reaches 76.83%, similar to Oort. Fig. 3 (b) presents results for EMNIST under IID partitioning. In this case, all four algorithms achieve final accuracies exceeding 80%. However, FCH-FL outperforms the others with the highest accuracy at 87.52%. This indicates that FCH-FL has a certain advantage in the IID partitioning of EMNIST and demonstrates improvements in model accuracy.

Heterogeneous data balancing evaluation. Our focus is to observe the impact of Non-IID data caused by long-tailed local data on global accuracy and the sensitivity of different algorithms to this type of heterogeneous data. We study the global model accuracy of different FL frameworks to verify whether the heterogeneous client balancing algorithm in FCH-FL can improve global model accuracy under such conditions. In the experiment, we set the client skewness level ($IF = 0.5$) and the ratio of clients with long-tailed data ($\mu = 0.5$). Fig. 4 illustrates the global accuracy performance of four algorithms across different rounds on CIFAR-10 and EMNIST.

As shown in Fig. 4 (a), in the Non-IID scenario caused by local long-tailed data in CIFAR-10, the global accuracy curve of the FedAvg exhibits significant oscillations compared to the IID scenario. This phenomenon arises due to the absence of any specific measures in FedAvg to address data heterogeneity, resulting

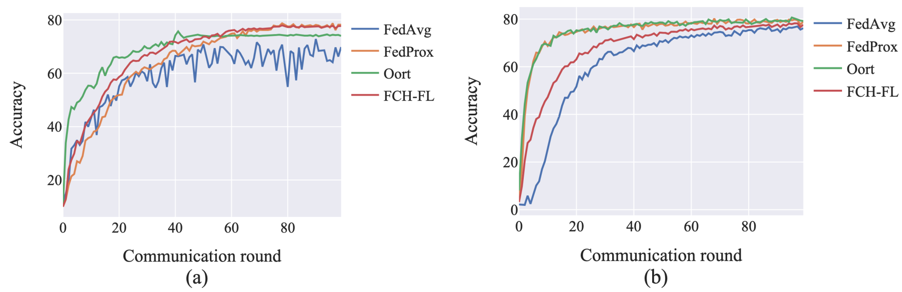

Fig. 4. Global model accuracy vs. training rounds for different FL frameworks with long-tailed local data ($IF = 0.5$, $\mu = 0.5$): (**a**) CIFAR-10 and (**b**) EMNIST.

in lower convergence performance and the lowest model accuracy in heterogeneous scenarios. Despite adopting different strategies, FedProx, Oort, and the FCH-FL algorithm all have a positive impact on client-local long-tailed data, resulting in higher final accuracy. Among them, the leader is FCH-FL, with an accuracy of 75.70% in the 100th round in the same scenario. This improvement of 5.89% compared to FedAvg is attributed to our proposed client heterogeneity balancing algorithm, which enhances the generalization ability of local models through global prior information and correction of local models.

We further validate the client balancing algorithm on a larger-scale EMNIST dataset, as shown in Fig. 4 (b). In the Non-IID scenario caused by long-tailed data from local clients, the best accuracy is achieved by Oort, followed by FedProx, with FCH-FL slightly behind. This is attributed to the increased number of clients, providing more opportunities for client selection strategies. However, compared to FedAvg in the same scenario, FCH-FL achieves 1.4% accuracy improvement. The results demonstrate the effectiveness of the FCH-FL algorithm in enhancing the final accuracy of heterogeneous FL with long-tailed local client data, as shown in Tab. 1. Additionally, the experiments indicate that our approach has a stronger advantage when the number of clients is relatively small. However, as a controlled experiment, we only examined the effectiveness of the client heterogeneity balancing algorithm in isolation. Combining it with our improved client selection strategy, FCH-FL is expected to achieve even better accuracy.

Table 1. Final accuracies on different datasets with IID and long-tail local data

Dataset	FedAvg	FedProx	Oort	FCH-FL
CIFAR-10 IID	74.13%	76.05%	**77.83%**	76.83%
CIFAR-10 Long-tail	69.81%	75.21%	74.30%	**75.70%**
EMNIST IID	83.91%	84.40%	86.40%	**87.52%**
EMNIST Long-tail	76.17%	78.98%	**79.15%**	77.57%

Client selection strategy evaluation. We focus on observing how the client selection strategy in FCH-FL impacts the improvement of global model accuracy. For CIFAR-10, we employ a Non-IID setting based on Dirichlet distribution [4] and distribute it among 50 clients. In each round, 20% clients were selected, totaling 200 training rounds. For EMNIST under the Non-IID scenario, we also used a distribution inspired by Dirichlet distribution. We split the dataset into 80% training and 20% testing, aligning with the Non-IID concept. During training, we divide the training set into 360 clients, with 10% of them participating in each round. Fig. 5 shows our comparison of the global model accuracy on different datasets.

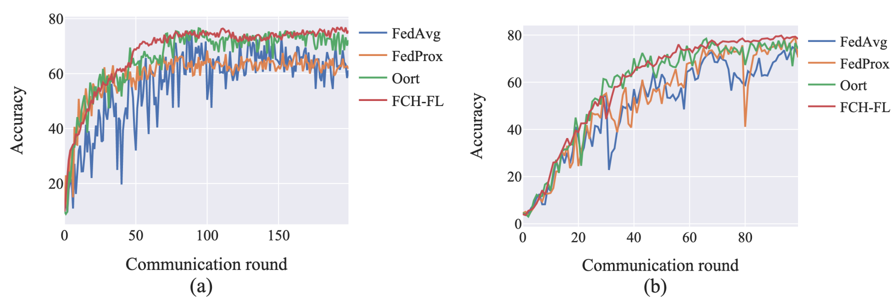

Fig. 5. Global model accuracy vs. training rounds on CIFAR-10 with skewed label distribution ($\beta = 0.2$) and EMNIST datasets: (**a**) CIFAR-10 and (**b**) EMNIST.

As depicted in Fig. 5 (a), in CIFAR-10, due to the lack of specific measures to address data heterogeneity, FedAvg exhibits severe accuracy oscillations and attains the lowest model accuracy upon convergence. In contrast, FedProx, Oort, and the FCH-FL exhibit better convergence performance and higher final accuracy with their distinct strategies for handling skewed label distributions. Particularly, the accuracy of FCH-FL at the 200th round is 75.63%. Compared to the IID scenario, this represents a mere 1.2% decline. Furthermore, it surpasses FedAvg by 14.42% and Oort by 3.46%. Validation results on EMNIST, as shown in Fig. 5 (b), reveal a decrease in accuracy for four algorithms in the Non-IID scenario created by skewed labels. However, FCH-FL consistently achieves the highest final accuracy, reaching 78.34%, which is 7.82% higher than FedAvg and 4.17% higher than Oort.

The results demonstrate that the FCH-FL exhibits better adaptability to scenarios with skewed label distributions resulting in Non-IID settings compared to baselines. This superiority arises from the finer-grained client selection in FCH-FL. Unlike other selection strategies, our method prioritizes accuracy. The utility function defined in our algorithm penalizes high-accuracy-contributing clients less for time delays, enabling them to participate more in each iteration, thus improving the global model's accuracy performance.

4.3 Efficiency evaluation

We study how system utility impacts end-to-end performance in the Straggler scenario and validate the effectiveness of FCH-FL in enhancing system heterogeneity. The research findings indicate that, in some cases, although both scenarios converge in the same number of rounds, the Straggler scenario takes slightly longer to reach the target accuracy (as shown in [7]). To model the presence of Stragglers in FL, we used Zipf's distribution ($\alpha = 1.2$) to simulate system heterogeneity and evaluate the time-to-accuracy metric, comparing it with the baselines. Figs. 6 (a) and (b) depict the experimental results for time-to-accuracy on CIFAR-10 and EMNIST. This indicates that Oort's efficiency has a slight advantage in reaching initial accuracy. However, at a certain convergence accuracy, FCH-FL surpasses Oort.

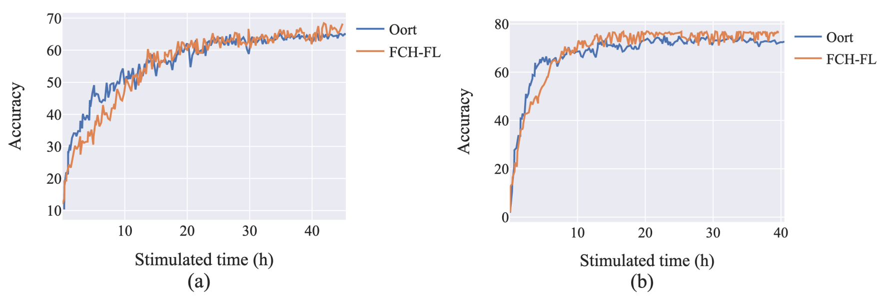

Fig. 6. Accuracy vs. training times for different FL frameworks with Non-IID and system heterogeneity: (**a**) CIFAR-10 and (**b**) EMNIST.

Tab. 2 presents the time comparison of FCH-FL and Oort to reach the same accuracy levels. We observe that on CIFAR-10, FCH-FL takes less time to reach 60% accuracy, and on EMNIST, it takes less time to reach 70% accuracy. This can be attributed to the fact that we balance local models for long-tailed local data, thereby improving the generalization performance of local models. Furthermore, we segment clients preserving those contributing high accuracy, which significantly enhances the convergence performance.

5 Related works

To address the issue of non-IID data, some studies have focused on improving the generalization of local models in FL, such as FedProx [9], FedNova [19], and SCAFFOLD [6]. These methods provide partial solutions to the non-IID problem from different perspectives. However, the performance of FL optimization methods can be compromised when dealing with a large number of clients.

Table 2. Time-to-accuracy of two algorithms on different datasets

Dataset Accuracy	CIFAR-10 Oort	 FCH-FL	EMNIST Oort	 FCH-FL
50%	**8.82h**	10.15h	**2.40h**	3.90h
60%	18.77h	**13.69h**	**3.76h**	5.86h
65%	34.94h	**24.09h**	6.49h	**6.32h**
70%	-	-	8.68h	**8.16h**
75%	-	-	22.80h	**12.80h**

Furthermore, it has been shown that involving all clients is unnecessary and impractical. Researchers have explored client selection techniques to optimize FL, considering only data or system heterogeneity, such as [13,20,24]. Some research works simultaneously address both data and system heterogeneity by adopting different selection criteria for clients, for instance [1, 18, 22, 27]. In these methods, asynchronous algorithms that consider both types of heterogeneity aim to improve statistical utility. However, these asynchronous approaches may not prioritize statistical utility due to their efficiency-oriented strategies [5,21,27]. While these methods consider both data and system heterogeneity, they introduce significant overhead in client selection. Oort [7] improves training performance by selecting clients with higher loss values and faster training speed, utilizing an MAB approach for efficient client selection. NCCB [14] establishes relationships between client features and rewards, enabling intelligent selection of client combinations. FedCG [26] proposes adaptive client selection and gradient compression to tackle the challenges. These methods address the challenges of considering both types of heterogeneity while mitigating additional overhead. However, they still have limitations in effectively addressing the poor generalization of local models caused by severe non-IID issues, such as local data label shifts. Our method is the first to consider both types of data heterogeneity simultaneously, improve model generalization ability in the local model training process, and refine the client selection strategy in the client selection process. The ablation experiment proves the effectiveness of our client selection strategy.

6 Conclusion

To address the challenges of data and system heterogeneity in FL, a fast convergence heterogeneous FL via active-passive collaboration is proposed. For local data heterogeneity, we propose a heterogeneity data balancing algorithm on the client side, which adjusts the Logit layer of the cross-entropy loss function using prior knowledge of the global data distribution. This adjustment increases the margin distance between different classes, thereby improving the generalization performance of local models. Recognizing the inconsistency between data quality and training time, we enhance the existing client utility function and propose a client selection algorithm for heterogeneous FL via MAB. It comprehensively considers and balances statistical and system utility, efficiently selecting suitable clients for FL training to improve the time-to-accuracy performance. We simulated heterogeneous scenarios on the CIFAR-10 and EMNIST datasets and compared the results with homogeneous scenarios to assess accuracy, convergence, and efficiency. The results showed that FCH-FL outperformed baseline methods in performance.

Declarations

Funding

This work was supported by the National Natural Science Foundation of China under Grant No. 62102074 and the Natural Science Foundation of Liaoning Province No. 2024-MSBA-49.

References

1. Chang, T., Li, L., Wu, M., Yu, W., Wang, X., Xu, C.: Pagroup: Privacy-aware grouping framework for high-performance federated learning. J. Parallel Distributed Comput. **175**, 37–50 (2023)
2. Cohen, G., Afshar, S., Tapson, J., Van Schaik, A.: Emnist: Extending mnist to handwritten letters. In: 2017 international joint conference on neural networks (IJCNN). pp. 2921–2926. IEEE (2017)
3. Fang, C., Guo, Y., Hu, Y., Ma, B., Feng, L., Yin, A.: Privacy-preserving and communication-efficient federated learning in internet of things. Computers & Security **103**, 102199 (2021)
4. Hsu, T.M.H., Qi, H., Brown, M.: Measuring the effects of non-identical data distribution for federated visual classification. arXiv preprint arXiv:1909.06335 (2019)
5. Jiang, Z., Wang, W., Li, B., Li, B.: Pisces: efficient federated learning via guided asynchronous training. In: Proceedings of the 13th Symposium on Cloud Computing. pp. 370–385 (2022)
6. Karimireddy, S.P., Kale, S., Mohri, M., Reddi, S., Stich, S., Suresh, A.T.: Scaffold: Stochastic controlled averaging for federated learning. In: International conference on machine learning. pp. 5132–5143. PMLR (2020)
7. Lai, F., Zhu, X., Madhyastha, H.V., Chowdhury, M.: Oort: Efficient federated learning via guided participant selection. In: 15th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 21). pp. 19–35 (2021)
8. Li, Q., Diao, Y., Chen, Q., He, B.: Federated learning on non-iid data silos: An experimental study. In: 2022 IEEE 38th International Conference on Data Engineering (ICDE). pp. 965–978. IEEE (2022)
9. Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V.: Federated optimization in heterogeneous networks. Proceedings of Machine learning and systems **2**, 429–450 (2020)
10. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Artificial intelligence and statistics. pp. 1273–1282. PMLR (2017)
11. Menon, A.K., Jayasumana, S., Rawat, A.S., Jain, H., Veit, A., Kumar, S.: Long-tail learning via logit adjustment. arXiv preprint arXiv:2007.07314 (2020)

12. Nguyen, D.C., Ding, M., Pathirana, P.N., Seneviratne, A., Li, J., Poor, H.V.: Federated learning for internet of things: A comprehensive survey. IEEE Communications Surveys & Tutorials **23**(3), 1622–1658 (2021)
13. Nishio, T., Yonetani, R.: Client selection for federated learning with heterogeneous resources in mobile edge. In: ICC 2019-2019 IEEE international conference on communications (ICC). pp. 1–7. IEEE (2019)
14. Pan, Q., Cao, H., Zhu, Y., Liu, J., Li, B.: Contextual client selection for efficient federated learning over edge devices. IEEE Trans. Mob. Comput. **23**(6), 6538–6548 (2024)
15. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H.R., Albarqouni, S., Bakas, S., Galtier, M.N., Landman, B.A., Maier-Hein, K., et al.: The future of digital health with federated learning. NPJ digital medicine **3**(1), 119 (2020)
16. Tan, J., Wang, C., Li, B., Li, Q., Ouyang, W., Yin, C., Yan, J.: Equalization loss for long-tailed object recognition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11662–11671 (2020)
17. Van Horn, G., Perona, P.: The devil is in the tails: Fine-grained classification in the wild. arXiv preprint arXiv:1709.01450 (2017)
18. Wang, H., Kaplan, Z., Niu, D., Li, B.: Optimizing federated learning on non-iid data with reinforcement learning. In: IEEE INFOCOM 2020-IEEE Conference on Computer Communications. pp. 1698–1707. IEEE (2020)
19. Wang, J., Liu, Q., Liang, H., Joshi, G., Poor, H.V.: Tackling the objective inconsistency problem in heterogeneous federated optimization. Advances in neural information processing systems **33**, 7611–7623 (2020)
20. Wu, H., Wang, P.: Node selection toward faster convergence for federated learning on non-iid data. IEEE Transactions on Network Science and Engineering **9**(5), 3099–3111 (2022)
21. Wu, Q., Chen, X., Ouyang, T., Zhou, Z., Zhang, X., Yang, S., Zhang, J.: Hiflash: Communication-efficient hierarchical federated learning with adaptive staleness control and heterogeneity-aware client-edge association. IEEE Transactions on Parallel and Distributed Systems **34**(5), 1560–1579 (2023)
22. Xin, F., Zhang, J., Luo, J., Dong, F.: Federated learning client selection mechanism under system and data heterogeneity. In: 2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD). pp. 1239–1244. IEEE (2022)
23. Yang, M., Guo, T., Zhu, T., Tjuawinata, I., Zhao, J., Lam, K.Y.: Local differential privacy and its applications: A comprehensive survey. Computer Standards & Interfaces p. 103827 (2023)
24. Yu, L., Albelaihi, R., Sun, X., Ansari, N., Devetsikiotis, M.: Jointly optimizing client selection and resource management in wireless federated learning for internet of things. IEEE Internet of Things Journal **9**(6), 4385–4395 (2021)
25. Yu, L., Albelaihi, R., Sun, X., Ansari, N., Devetsikiotis, M.: Jointly optimizing client selection and resource management in wireless federated learning for internet of things. IEEE Internet Things J. **9**(6), 4385–4395 (2022)
26. Zhai, R., Jin, H., Gong, W., Lu, K., Liu, Y., Song, Y., Yu, J.: Adaptive client selection and model aggregation for heterogeneous federated learning. Multim. Syst. **30**(4), 211 (2024)
27. Zhang, J., Wang, J., Li, Y., Xin, F., Dong, F., Luo, J., Wu, Z.: Addressing heterogeneity in federated learning with client selection via submodular optimization. ACM Trans. Sens. Networks **20**(2), 48:1–48:32 (2024)