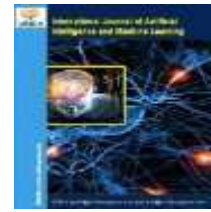




International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

Joint Channel and Learning Quality-Aware Client Scheduling in Wireless Federated Learning

Dr. Hemant A. Wani¹, Dr. Anil L. Wanare², Dr. Chandrashekhar Patil³, Dr. Neelima R. Kolhare⁴, Dr. Yogita D. Kapse⁵, Shailendra Anant Hingonekar⁶

¹Associate Professor, E&TC Engg. JSPM's BSIoT, Pune hawani_entc@jspmbsiotr.edu.in

²Professor, E&TC Engg. JSPM's Group of Institutes Eudoxia Research Centre, Guwahati, India anillaxman369@gmail.com

³Associate Professor, E&TC Engg. JSPM's BSIoT, Pune cspatil_entc@jspmbsiotr.edu.in

⁴Associate Professor, COEP Tech. University, Pune nrk.extc@coeptech.ac.in

⁵Assistant Professor, COEP Tech. University, Pune ydk.extc@coeptech.ac.in

⁶Assistant Professor, E&TC Engg. SSVPS Bapusaheb Shivajirao Deore College of Engineering, Dhule (Maharashtra) shailendra.hingonekar@gmail.com

Abstract

Federated learning (FL) has emerged as a promising distributed learning paradigm for Internet of Things (IoT) applications, enabling participating clients to train local models using their private data and periodically upload model updates for global aggregation. By keeping raw data at the client side, FL enhances data privacy while reducing the need for centralized data collection. However, the limited and heterogeneous nature of wireless communication resources poses significant challenges to efficient client participation and model training. In this letter, we propose a client scheduling strategy for wireless federated learning that jointly considers channel quality and learning quality to identify clients that can contribute more effectively to the global training process. Unlike conventional scheduling approaches that primarily prioritize clients according to their wireless channel conditions, the proposed strategy incorporates both communication and learning-related characteristics in the client selection process. We evaluate the proposed approach against conventional channel-quality-based scheduling methods. Experimental results demonstrate that the proposed strategy substantially improves the FL training performance, achieving higher model accuracy and faster convergence while effectively utilizing the available wireless resources.

Keyword: Federated learning (FL), Internet of Things (IoT), wireless federated learning, client scheduling, channel quality, learning quality.

Introduction

In recent years, the rapid development of the Internet of Things (IoT), big data, and artificial intelligence (AI) has enabled intelligent services to permeate various aspects of modern life. In contrast to conventional centralized machine learning (ML) frameworks, deploying AI at the network edge by leveraging data generated from a large number of mobile and IoT devices has emerged as a promising approach for enabling intelligent wireless networks and supporting future 6G systems [1]. However, the distributed nature of data and computation in such networks introduces several critical challenges. First, transmitting raw data from mobile devices to a centralized server for model training may expose users' sensitive information and consequently raise significant privacy concerns [2]. Second, the large volume of data generated by massive numbers of devices can impose substantial communication overhead, resulting in increased transmission latency and potentially degrading the overall learning performance. Federated learning (FL) has emerged as an effective solution to address these challenges [3]. In FL, participating devices collaboratively train a global model by performing local computations on their own datasets and transmitting only model updates to a central server, thereby avoiding the direct exchange of raw data. By exploiting the distributed computing and storage capabilities of massive mobile and IoT devices, FL enables collaborative model training while preserving data locality. Nevertheless, achieving a target level of global model accuracy generally requires multiple rounds of interactions between the central server and participating clients. Therefore, the time required for an FL system to achieve the desired model accuracy is an important metric for evaluating its learning efficiency and overall performance [4].

The training time of an FL system is primarily determined by two components: local computation time and communication time. In wireless FL systems, the latter can become a significant bottleneck because wireless resources are limited and channel conditions are time-varying. In particular, communication delays arising from dynamic wireless environments can substantially affect the efficiency and convergence performance of FL systems

[5], [6]. Consequently, appropriate client selection and scheduling mechanisms are essential for efficiently utilizing scarce wireless resources and reducing the time required for model training.

In addition to communication conditions, the heterogeneous nature of client data is another important factor that influences FL performance. Since clients are typically distributed across different physical environments and generate data according to their individual usage patterns, the data distributions across clients are generally non-identically distributed (non-IID) [7]. As a result, different clients may contribute unequally to the global learning process. One intuitive approach to assessing a client's contribution is to consider the quantity of its local data. For instance, a client possessing a larger number of training samples may provide more information to the global model [11]. However, the number of local samples alone may not fully characterize the learning value of a client. The characteristics of local model updates provide another means of evaluating client importance. In particular, the magnitude of the local gradient can reflect the degree to which a client's local model differs from the current global model. A larger gradient may indicate that the client's local data contain information that is more distinct from the knowledge already captured by the global model [8]–[10]. Therefore, prioritizing clients that provide highly informative local updates can potentially improve the learning efficiency and convergence performance of an FL system.

These observations indicate that effective client scheduling in wireless FL should account for both communication conditions and learning contributions. Scheduling clients solely according to their channel quality may improve communication efficiency but may overlook clients whose local data or model updates are particularly valuable for global learning. Conversely, selecting clients solely according to their learning contribution may result in excessive communication delays or inefficient utilization of wireless resources. Motivated by this tradeoff, this work investigates a client scheduling strategy that jointly considers channel quality and learning quality to improve the overall training efficiency of wireless FL systems.

The main contribution of this work is a joint channel- and learning-quality-based client scheduling strategy for wireless FL. By considering both the communication conditions of clients and the informativeness of their local learning updates, the proposed approach aims to select clients that provide an effective balance between communication efficiency and learning contribution. The effectiveness of the proposed strategy is evaluated through experiments and compared with conventional client scheduling methods that primarily consider channel quality. The results demonstrate that jointly considering communication and learning quality can improve model accuracy and accelerate the convergence of the FL training process.

Therefore, a fundamental challenge in wireless FL systems is to efficiently select clients that can contribute to the learning process while operating under limited wireless resources. To address this challenge, this letter proposes a client scheduling strategy based on a joint quality metric that captures both the communication and learning characteristics of participating clients. Specifically, the proposed metric jointly considers channel quality and learning quality to determine the participation priority of each client in a given training round. By incorporating both factors into the scheduling decision, the proposed approach aims to achieve a more effective tradeoff between communication efficiency and learning contribution. Simulation results demonstrate that the proposed scheduling strategy can improve model accuracy and accelerate convergence compared with conventional channel-quality-based client selection methods.

The main contributions of this letter are summarized as we propose a novel learning-quality metric based on gradient information entropy and combine it with channel quality to form a joint client-quality metric. Using this metric, a low-complexity quality-aware scheduling strategy is developed to prioritize clients according to both communication and learning conditions. Simulation results show that the proposed approach improves model accuracy and convergence speed, particularly under limited wireless resources.

System Model

In this section, we present the system model considered for wireless federated learning (FL). We first describe the wireless communication model and characterize the channel conditions between the server and participating clients. We then introduce the learning model, including the local training process and the formulation of client learning quality. Finally, we describe the overall FL training procedure and the interaction between the communication and learning processes.

If your next subsections are specifically A. Wireless Communication Model, B. Learning Model, and C. Federated Learning Process, this introduction provides a clean transition into them.

A. Wireless Communication Model (Channel Model) : We consider a wireless network comprising a single-antenna access point (AP) and N single-antenna clients. The total system bandwidth B is equally divided into M orthogonal sub-channels, each having a bandwidth of B/M . In each communication round, a sub-channel is exclusively assigned to one participating client. Due to limited spectrum resources and possible sub-channel unavailability, only K clients can participate in each round. Hence, the number of scheduled clients satisfies

$$K \leq M < N.$$

Let $d_k(t)$ denote the distance between client k and the AP located at the centre of the cell during communication round t . The channel gain of client k over sub-channel m , denoted by $H_{k,m}(t)$, accounts for both large-scale path loss and small-scale fading. Accordingly, the achievable uplink data rate for client k on sub-channel m , while transmitting its local model to the AP, is denoted by $r_{k,m}(t)$ and depends on the corresponding signal-to-noise ratio (SNR), $\gamma_{k,m}(t)$.

B. Learning Model

We consider a federated learning (FL) network in which multiple clients collaboratively train a global model under the coordination of an access point (AP). Each client possesses a local dataset containing labelled samples. Due to limited communication resources, only a subset of clients can upload their locally trained model parameters or updates to the AP and participate in global model aggregation during each communication round. At the beginning of each training round, the AP broadcasts the current global model to all clients. Each client independently trains the received global model using its local dataset. After local training, the AP selects a subset of clients to upload their local model updates for aggregation and subsequent global model updating. Let K clients be selected to participate in the t -th communication round. For client k , its local dataset is denoted by

$$D_k = \{(s_k^1, y_k^1), (s_k^2, y_k^2), \dots, (s_k^{s_k}, y_k^{s_k})\} \dots \dots (1)$$

where s_k is the size of the local data set. Here, the distribution of clients' data is assumed to be non-independent and identical distribution (non IID) because most clients may only obtain certain types of samples based on their environment or activity characteristics.

Let $L(w_k(t))$ be the average local loss function value of client k in round t , where $W_k(t)$ is the vector of the parameters of the local model, x_k is the data sample, and $\ell()$ is the loss function. Then the average local loss function value of client k can be expressed as

$$L(W_k(t)) = \frac{1}{s_k} \sum_{i=1}^{s_k} \ell(W_k(t), x_k^i, y_k^i) \dots \dots (2)$$

the k selected clients, the global loss function is formulated as a weighted average of the local loss functions, where the weight assigned to each client is proportional to the size of its local dataset:

$$L(W) = \frac{1}{\sum_{k=1}^K s_k} \sum_{k=1}^K s_k L(w_k(t)) \dots \dots (3)$$

Here, $W^{(t)}$ denotes the global model parameter vector at communication round t . The weighted formulation ensures that clients with larger local datasets contribute proportionally more to the global objective.

For client k , the gradient vector computed using its local dataset during the t -th communication round is given by

$$\begin{aligned} g_k(t) &= \nabla L_k(w_k(t)) \\ &= \sum_{i=1}^{s_k} \frac{\partial \ell(w_k(t), x_k^i, y_k^i)}{\partial w_k(t)} \dots \dots (4) \end{aligned}$$

The local gradients or model updates are subsequently transmitted by the selected clients to the AP, which aggregates the received updates to obtain the updated global model. This process is repeated over multiple communication rounds until the global model converges or a predefined training criterion is satisfied. Assuming that the current communication round is t , the local model parameters of client k for the subsequent communication round $t + 1$ can be updated using the gradient descent rule as follows:

$$w_k(t+1) = w_k(t) - \eta_k \frac{1}{K} \sum_{k=1}^K g_k(t) \dots \dots (5)$$

where $w_k(t)$ and $w_k(t+1)$ denote the local model parameter vectors of client k at communication rounds t and $t + 1$, respectively; $g_k(t) = \nabla L_k(w_k(t))$ represents the gradient of the local loss function; and η_k denotes the learning rate of client k .

C. Federated Learning Process

Federated Learning (FL) is a distributed machine learning paradigm in which multiple clients collaboratively train a global model without sharing their raw local data. In each communication round, the central server initializes

and distributes the current global model to selected clients. Each client independently trains the model using its private local dataset and computes updated model parameters. The clients transmit only the learned model updates or parameters to the central server, preserving data locality and privacy. The server aggregates the received updates using an appropriate strategy, such as FedAvg, to generate an improved global model for the next training round.

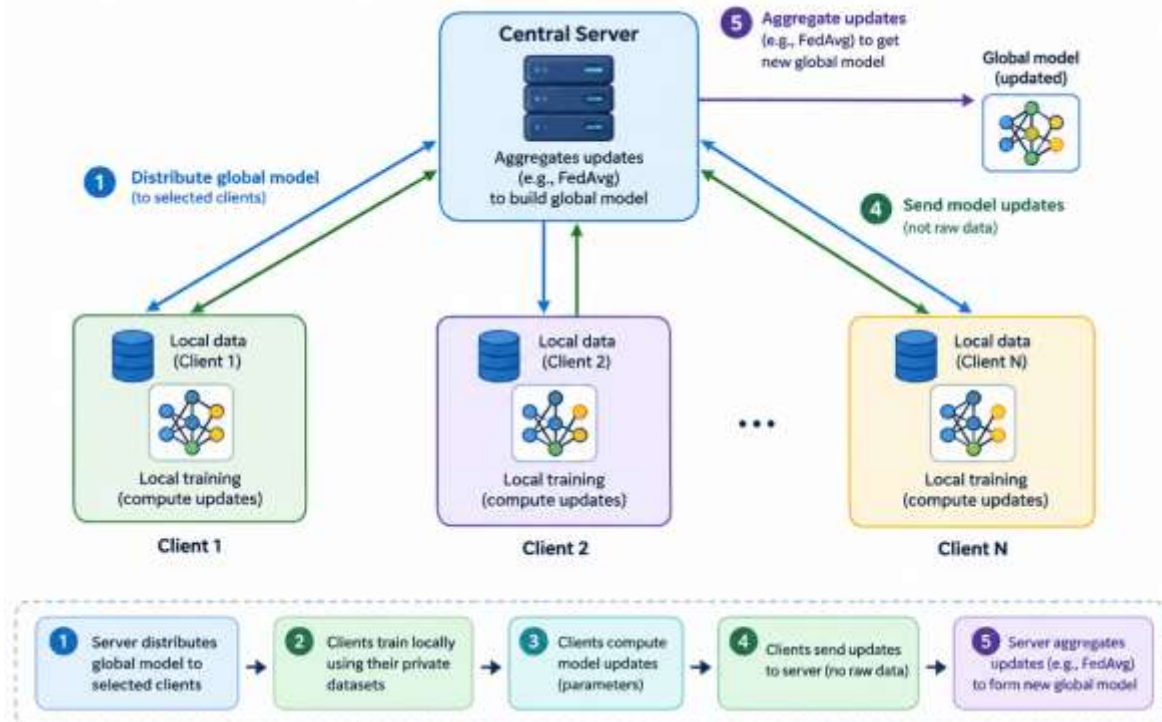


Fig. 1. Overview of the federated learning framework with local client training, model update transmission, and global model aggregation at the central server.

III. Optimization Of User Scheduling In A Wireless FI System

In this section, we first define the channel quality and learning quality of each client participating in the wireless federated learning (FL) system. Based on these two factors, we then introduce a combined quality metric to jointly evaluate the communication and learning contributions of the clients. Finally, a wireless resource scheduling method is proposed based on the combined quality metric, with the objective of efficiently selecting and allocating wireless resources to clients while improving the overall FL performance and reducing communication overhead.

A. Channel Quality

In this work, the system adopts a synchronous model upload mechanism. During each round of local model transmission, a fixed time window T_{up} is allocated to each client to complete both the local model upload and channel estimation processes. Since the wireless system has limited communication resources B the channel conditions of individual clients directly influence the resource scheduling cost. In particular, clients experiencing poor channel conditions require a larger share of wireless resources to achieve reliable and efficient communication. We assume that an Orthogonal Frequency Division Multiple Access (OFDMA) scheme is employed for transmitting the local models. Due to the complex and dynamic nature of wireless environments, wireless channels generally exhibit frequency-selective fading. Therefore, the frequency-selective channel responses of the clients are considered as an important indicator of channel quality.

Let the channel frequency response of client k on sub-channel m be denoted by $H_k(f_m)$, where f_m represents the center frequency of sub-channel m . The system collects the channel responses of all clients across all available sub-channels to construct a channel quality matrix Q_{chan} , which is used to characterize the current channel conditions of the participating clients. Accordingly, the channel quality of client k on sub-channel m is expressed as follows:

$$Q_{chan}(k, m) = H_k(f_m), k \in K \text{ and } m \in M \dots \dots \dots (6)$$

The channel quality matrix Q_{chan} is updated at the beginning of each training round to capture the time-varying wireless channel conditions as the federated learning process progresses.

B. Learning Quality

To improve the convergence of the global model, it is important to evaluate how much each client contributes to the model update in every training round. In machine learning, the gradient indicates the direction in which the model parameters should be changed in the next update. According to information theory, the entropy of a random variable represents the average amount of information contained in its possible outcomes. Therefore, we assume that the elements of the gradient vector follow a certain probability distribution and use their entropy to measure the amount of useful convergence information contained in the local model parameters [12].

A higher gradient information entropy indicates that the gradient values are more evenly distributed. This means that the model parameters are being adjusted in different directions, providing more diverse information for model training. On the other hand, when the gradient values are concentrated within a small numerical range, the gradient information entropy is lower. This indicates that the parameter updates are less diverse and provide less information for improving the model. The gradient information entropy of client k , denoted by E_k , is defined as follows:

$$E_k = - \sum_{g \in g_k} g \log_2(g) \dots \dots \dots (7)$$

On the other hand, in machine learning training, the loss value indicates how much the local model needs to be improved. Therefore, it can be used to represent the importance of a client's local model update to the convergence of the global model. In general, the product of two factors can be used to represent their combined effect. In this case, the gradient entropy represents the amount of useful information, while the average local loss value represents the importance of the model update. Therefore, we use the product of these two factors to define the learning quality index Q_{learn} .

The learning quality index of client k is given by the product of its gradient information entropy and average local loss value.

$$Q_{learn}[k] = |E_k|L(w_k) \dots \dots \dots (8)$$

where E_k is the gradient information entropy of client k , and $L(w_k)$ is the average local loss value of client k . The corresponding values are updated in each training round to reflect the dynamic changes during the learning process. Since clients may operate under diverse environments and possess heterogeneous data distributions, we assume that the class distribution of local data is non-uniform across clients. Therefore, a local model that exhibits a greater deviation from the global model is considered to have higher learning value, as such differences may indicate the presence of unique or underrepresented data classes that have not yet been adequately captured by the global model.

B. Combined Quality

Both channel quality and learning quality should be considered when scheduling clients for communication. If the system gives too much priority to channel quality, it may select clients with good communication conditions but low learning contributions. As a result, the server may receive a large number of local models that provide limited benefits to global model convergence. On the other hand, if the system focuses mainly on learning quality, clients with poor channel conditions may be selected more frequently. Due to the limited wireless communication resources, this may increase the communication cost and reduce the number of local models that can be successfully uploaded in each training round. Therefore, a combined quality metric that considers both channel conditions and learning contributions is required for effective client scheduling. This metric can help the system achieve a better balance between communication efficiency and model convergence. Therefore, we propose a combined quality metric that jointly considers channel and learning qualities to balance communication efficiency and global model convergence. Since the relative importance of these factors varies throughout the training process, their weighting factors are dynamically adjusted. In the early training stage, greater emphasis is placed on channel quality to maximize client participation and accelerate initial model convergence. As training progresses, the weighting of learning quality is gradually increased to prioritize clients with greater contributions to global model convergence.

The combined quality metric denoted by Q_{comb} of client k on sub-channel m is proposed to take the following form:

$$Q_{comb}[k, m] = (\rho^{-t}(1 - \beta) + \beta)Q_{chan}[k, m] + (1 - (\rho^{-t}(1 - \beta) + \beta))Q_{learn}[k] \dots \dots (9)$$

where ρ is the decaying factor used to adjust the changing trend of channel quality's weight, and β is the base weight of channel and learning qualities

D. Wireless Resource Scheduling

Because the total system bandwidth and the number of available sub-channels are limited, efficient allocation of wireless resources among clients is essential. For all candidate clients within the considered scenario, a value matrix is constructed based on the combined quality metrics and is used as the reference for wireless resource scheduling.

For client k , let $P_k(t)$ denote the maximum packet size that can be transmitted to the access point (AP) during communication round t . To efficiently allocate the available sub-channels, we propose a greedy resource scheduling algorithm. Specifically, the scheduler first identifies the largest combined quality metric reported by the candidate clients and allocates the corresponding sub-channel to the client associated with that metric. If the allocated sub-channel is insufficient to provide the throughput required for transmitting the client's model update, additional sub-channels are iteratively assigned to that client until its transmission requirement is satisfied. Once a client has been allocated a sufficient number of sub-channels, the corresponding client and allocated sub-channels are removed from the scheduling pool to prevent further allocation during the same scheduling round. The above procedure is repeated iteratively by selecting the largest remaining combined quality metric and allocating the corresponding sub-channel to the associated client. The scheduling process terminates when either all available sub-channels have been allocated or no further feasible allocation exists. This greedy strategy prioritizes clients with higher combined channel and learning quality while ensuring that sufficient wireless resources are assigned to support their model-upload requirements.

Experimental Results

In this section, we conduct extensive experiments to evaluate the effectiveness of the proposed wireless resource scheduling scheme in a federated learning (FL) environment. The evaluation focuses on the impact of the proposed scheme on client selection, wireless resource utilization, achievable throughput, communication latency, FL convergence, and global model learning accuracy. The proposed approach is evaluated under different wireless channel conditions and system configurations to examine its robustness and scalability. In particular, we investigate how the joint consideration of channel quality and learning-related information influences the selection and scheduling of participating clients. The achievable throughput and model-upload latency are analysed to quantify the communication efficiency of the proposed resource allocation strategy. Furthermore, the convergence behavior of the FL training process is evaluated in terms of the number of communication rounds required to achieve a target accuracy. The final global model accuracy is also compared with that of benchmark scheduling schemes to demonstrate the learning performance of the proposed approach.

For a comprehensive evaluation, the proposed scheme is compared with representative baseline resource allocation and client scheduling methods under identical experimental settings. The experimental results demonstrate that the proposed scheme can effectively exploit the available wireless resources, prioritize suitable clients, improve model-upload efficiency, and reduce communication latency. More importantly, the joint optimization of communication and learning quality leads to faster FL convergence and improved global model accuracy, highlighting the effectiveness of the proposed scheduling strategy for wireless federated learning systems.

A. Experimental Settings and Benchmark Schemes

In the experiments, we consider a single urban microcell with a radius of 600 m, where the base station (BS) is located at the centre of the cell. A total of 30 clients are uniformly distributed within the coverage area. The wireless channel between each client and the BS is modelled according to the Spatial Channel Model (SCM) [13], while the path loss is characterized using the COST231 propagation model. The average mobility speed of the clients is set to 60 km/h. The total system bandwidth is 10 MHz and is equally divided into 10 orthogonal sub-channels. For the federated learning (FL) task, a Multi-Layer Perceptron (MLP) is used as the global model, trained on the MNIST handwritten digit dataset. To study the impact of data imbalance, the training data are distributed in a non-IID manner by randomly partitioning them into shards, with each shard containing 120 images from a single digit class. The shards are randomly assigned to clients, creating heterogeneous local data distributions. Each client performs local training using Stochastic Gradient Descent (SGD) with a learning rate of 0.0001. All clients are assumed to have identical dataset sizes and computational capabilities so that performance differences mainly arise from data and wireless-channel heterogeneity. Four benchmark client-selection schemes are considered: Random Selection (RS), which randomly orders clients and assigns them to available sub-channels; Channel Quality Only (CQO), which prioritizes clients with better instantaneous channel conditions; Learning Quality Only (LQO), which selects clients based on learning quality while randomly assigning sub-channels; and All Users Uploading (AU), where all clients participate in aggregation regardless of channel conditions or resource

availability. The proposed method jointly considers channel and learning quality through the combined quality metric for resource-aware client selection and sub-channel allocation. All methods are evaluated under identical FL models, datasets, learning parameters, wireless conditions, and client populations for a fair comparison.

B. Performance Comparison

For a network comprising thirty clients and ten sub-channels, each client is assigned five data shards. All methods are evaluated under identical wireless and learning conditions. As shown in Fig. 2, the proposed combined quality indicator (8) is compared with the gradient norm [8], gradient divergence [9], and other client scheduling methods under both independent and identically distributed (IID) and non-IID data settings. The results demonstrate that the proposed method achieves superior performance, particularly in the non-IID setting, which can be attributed to the effectiveness of the proposed combined quality indicator.

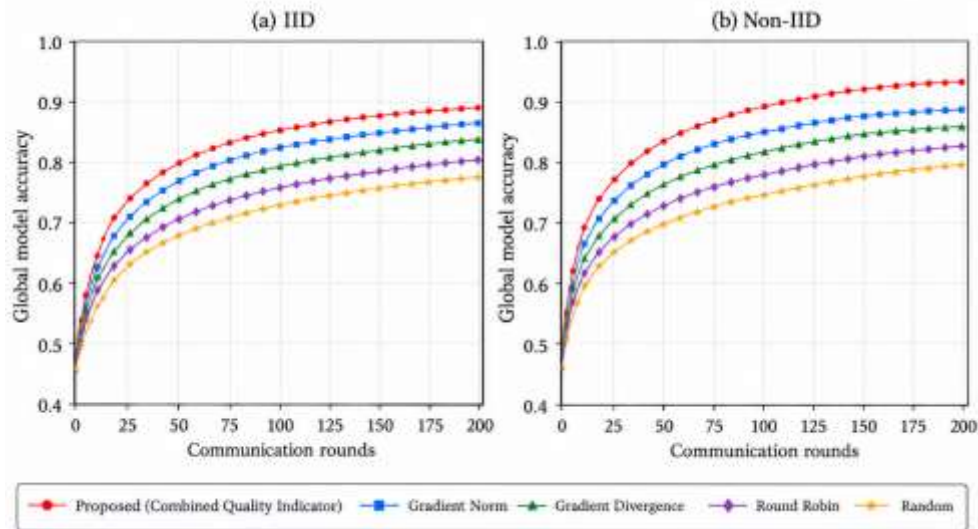


Fig. 2. Performance comparison of the proposed combined quality indicator with existing client scheduling methods under IID and non-IID data distributions.

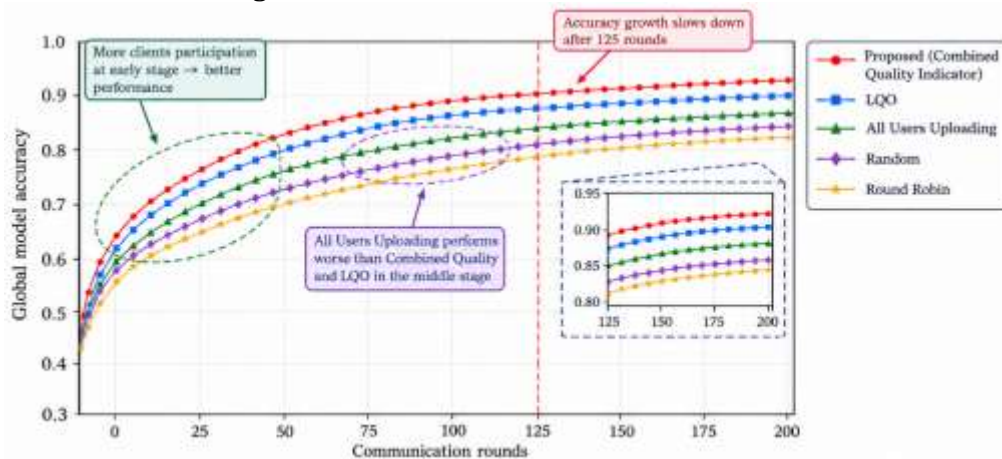


Fig. 3. Impact of limited local data on convergence and performance under restricted wireless resources.

In Fig. 3, the impact of the local data amount on system performance is evaluated by reducing the number of images per data shard to 30, reflecting the data limitations of edge clients. Methods involving more clients achieve better performance during the early training stage; however, their accuracy improvement slows significantly after 125 communication rounds. This behavior results from channel-based scheduling, which tends to repeatedly select a limited subset of clients, leading to insufficient data diversity in the later training stage. In the middle stage, All Users Uploading performs worse than the proposed Combined Quality and LQO methods because equal-weight aggregation allows low-quality local models to dilute the contributions of high-quality models. Overall, the proposed Combined Quality method achieves faster convergence and better performance under limited wireless resources.

In Table I, we evaluate the impact of different values of ρ and β on system performance. The results show that these balance factors affect the convergence performance to some extent. Based on the experimental results, we set $\rho = 1.015$ and $\beta = 0.5$. Due to the complex interactions among channel conditions, learning quality, and client participation, these values provide a suitable balance between communication efficiency and learning performance.

Table I. Impact of ρ and β on system accuracy

Balance Factors	40 Rounds	80 Rounds	120 Rounds	160 Rounds	200 Rounds
$\rho=1.10, \beta=0.0$	41.38	65.36	72.78	76.26	77.89
$\rho=1.07, \beta=0.0$	41.61	65.39	73.39	76.66	78.57
$\rho=1.05, \beta=0.0$	40.75	65.22	72.89	75.99	78.47
$\rho=1.015, \beta=0.0$	43.02	66.69	73.73	76.88	79.15
$\rho=1.015, \beta=0.3$	41.88	67.74	74.85	77.88	80.20
$\rho=1.015, \beta=0.5$	43.12	67.99	74.71	78.37	80.15
$\rho=1.015, \beta=0.7$	42.70	68.36	74.52	78.08	80.19

Since these factors affect system performance, determining their optimal values is challenging. In future work, we will investigate real-time optimization of the balance factors using reinforcement learning techniques.

Conclusion

This paper presented a quality-aware client scheduling approach to improve the training efficiency of federated learning in wireless networks. The proposed method jointly considers clients' wireless channel conditions and learning quality to determine client participation and resource allocation. Experimental results demonstrate that the proposed approach achieves faster convergence and improved FL performance compared with conventional scheduling schemes, particularly under limited wireless resources.

References

1. C. She et al., "A tutorial on ultrareliable and low-latency communications in 6G: Integrating domain knowledge into deep learning," *Proc. IEEE*, vol. 109, no. 3, pp. 204–246, Mar. 2021.
2. K. Wei et al., "Federated learning with differential privacy: Algorithms and performance analysis," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 3454–3469, 2020.
3. Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, pp. 1–19, 2019.
4. G. Zhu, Y. Wang, and K. Huang, "Broadband analog aggregation for low-latency federated edge learning," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 491–506, Jan. 2020.
5. H. H. Yang, Z. Liu, T. Q. S. Quek, and H. V. Poor, "Scheduling policies for federated learning in wireless networks," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 317–333, Jan. 2020.
6. M. M. Amiri and D. Gündüz, "Federated learning over wireless fading channels," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3546–3557, May 2020.
7. Y. Zhao, M. Li, L. Lai, N. Suda, D. Civan, and V. Chandra, "Federated learning with non-IID data," *arXiv:1806.00582*, 2018.
8. Y. He, J. Ren, G. Yu, and J. Yuan, "Importance-aware data selection and resource allocation in federated edge learning system," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13593–13605, Nov. 2020.
9. J. Ren, Y. He, D. Wen, G. Yu, K. Huang, and D. Guo, "Scheduling for cellular federated edge learning with importance and channel awareness," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7690–7703, Nov. 2020.
10. H. Guo, A. Liu, and V. K. N. Lau, "Analog gradient aggregation for federated learning over wireless networks: Customized design and convergence analysis," *IEEE Internet Things J.*, vol. 8, no. 1, pp. 197–210, Jan. 2021.
11. W. Xia, T. Q. S. Quek, K. Guo, W. Wen, H. H. Yang, and H. Zhu, "Multi-armed bandit-based client scheduling for federated learning," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7108–7123, Nov. 2020.
12. J. Ding, R. Calderbank, and V. Tarokh, "Gradient information for representation and modeling," in *Advances in Neural Information Processing Systems*, vol. 32. Red Hook, NY, USA: Curran Associates, 2019, pp. 1–10.
13. D. S. Baum, J. Hansen, J. Salo, G. D. Galdo, M. Milojevic, and P. Kyösti, "An interim channel model for beyond-3G systems: Extending the 3GPP spatial channel model (SCM)," in *Proc. IEEE 61st Veh. Technol. Conf. (VTC)*, vol. 5, 2005, pp. 3132–3136.