applied sciences

MDPI

Article

Joint Client and Resource Optimization for Federated Learning in Wireless IoT Networks

Jie Zhao [1,2], Yiyang Ni [2,3,*] and Yulun Cheng [3]

1 College of Physics and Information Engineering, Jiangsu Second Normal University, Nanjing 210013, China; zhaojie@jssnu.edu.cn
2 Jiangsu Province Engineering Research Center of Basic Education Big Data Application, Jiangsu Second Normal University, Nanjing 210013, China
3 Jiangsu Key Laboratory of Wireless Communications, Nanjing University of Posts and Telecommunications, Nanjing 210003, China; chengyuluen@163.com
* Correspondence: niyy@jssnu.edu.cn

Abstract: Federated learning (FL) is a promising technique to provide intelligent services for the internet of things (IoT). By transmitting the model parameters instead of user data between the client and central server, FL greatly improves the user privacy and reduces transmission latency. However, due to the fading effects of the wireless channel, the outage of wireless transmission degenerates the learning efficiency when FL is applied in wireless IoT networks. In order to address this issue, we investigate the joint optimization of client selection and wireless resource allocation in FL-aided cellular IoT networks. By taking both the amount of training data and wireless resource consumption into consideration, we formulate the problem as a mixed integer non-linear programming to maximize the utility of the network. To solve the problem effectively, an alternative direction-based algorithm is proposed by decomposing the original problem into two sub problems. The simulation results indicate that the proposed algorithm substantially improves the FL learning performance and reduces the consumption of wireless resources compared with existing methods.

Keywords: federated learning; client selection; resource allocation; internet of things

Citation: Zhao, J.; Ni, Y.; Cheng, Y. Joint Client and Resource Optimization for Federated Learning in Wireless IoT Networks. *Appl. Sci.* **2024**, *14*, 542. https://doi.org/10.3390/app14020542

Academic Editor: Christos Bouras

Received: 6 November 2023
Revised: 24 December 2023
Accepted: 5 January 2024
Published: 8 January 2024

1. Introduction

With the rapid development of artificial intelligence (AI), massive intelligent devices have been deployed in the wireless internet of things (IoT), which makes the ability of providing intelligent services one of the most important evolutionary directions for future IoT. As a basic element of AI, machine learning (ML) [1] is widely utilized by various intelligent applications. However, the centralized training manner in ML becomes more and more inefficient and data transmission and privacy problems arise because massive amounts of training data are generated and stored in various intelligent devices other than the central server. Federated learning (FL) [2,3] is a desired solution to handle this mismatch. As a distributed ML, FL only requires the transmission of model parameters other than the data themselves between client and central server, which largely reduces the amount of data transmission. Meanwhile, the user's privacy is also well protected, since the data avoid being transmitted to the central server.

Despite the advantage of FL, the communication issue should be addressed when deploying it in wireless IoT networks [4]. Specifically, because of the fading effects of the wireless channel, the transmission of model parameters may experience an outage, which degenerates the performance of FL [5]. Towards this issue, an admission control algorithm is proposed in [6], where the number of accessed devices is considered as the optimization target. Furthermore, in [7], the quality of local training and the channel state are utilized to decide the accessed devices. In addition to admission control, the wireless resource optimization is also efficient for FL-aided IoT networks. For example, in [8], a joint optimization algorithm is proposed to improve the communication reliability of FL-aided wireless networks, where the client selection and power allocation are employed to reduce the loss of the trained model. In [9], split learning and FL are combined to handle the diversity of clients with different channel states and computational capabilities. In this algorithm, client selection is achieved by a multi-arm bandit scheme, which employs both the channel states and local model as the optimized objective. In [10], a blockchain-based FL is proposed for wireless computing power networks, where the client selection is achieved by a evolutionary game-based incentive scheme. The incentive function takes the resource and security into account; however, the amount of local training data is not involved. Additionally targeting wireless computing power networks, a resource-aware FL is proposed in [11] that aims to reduce the energy consumption. The algorithm is employed to adjust the depth of the neural network and total training round without involving power and wireless channel selection. To handle the dynamics of wireless channels and network resources, in [12], the global FL models received in previous training are reused to replace erroneous local models. In this algorithm, the client selection is adopted by minimizing the accuracy loss in training data, so it focuses on the repair other than selecting superior wireless channels. All these previous works adopt the synchronous model; conversely, in [13], an asynchronous FL framework with client selection is proposed. In their optimization, the client availability and long-term fairness are taken into consideration to minimize latency. Lyapunov optimization is employed to tackle the asynchronous problem in an online manner. However, the amount of local training data is not considered in client selection. In [14], the training data of local model and wireless channel quality are jointly considered for asynchronous FL. The optimization objective is to reduce the variance and bias of the aggregated model updates, while the amount of local training data is also not involved. In [15], a joint optimization of bandwidth allocation and client scheduling is considered to achieve the ideal trade-off between training accuracy and latency. To solve the problem efficiently, the reformulation and decoupling are adopted, and the optimal resource allocation can be achieved by using an online algorithm. Nevertheless, the amount of local training data is not considered in the client scheduling. Another joint optimization of client selection and resource allocation for wireless FL is studied in [16], where the target is to maximize the total average number of active clients and transmission time. The Lyapunov optimization is employed to achieve an online-manner solution. Similarly, in [17], a joint optimization of client scheduling and resource allocation for hierarchical FL is investigated. The formulation simultaneously captures the uncertainty of the wireless channel and the weight gradient. However, in these former works, the joint optimization mainly focuses on the wireless channel and resource allocation without considering the amount of local training data in each client. In [18], both the bandwidth and power allocation are considered in the wireless resource optimization, and the objective is to maximize the number of accessed clients. However, in these existing works, the objective functions treat all nodes equally, and do not involve the amount of the training data in the client, which is worth being considered, because, in practice, the amount of collected data for each IoT node varies.

Motivated by these observations, we investigate the joint optimization of client selection and wireless resource allocation in FL-aided IoT networks, and the major contributions are as follows.

- We developed a joint optimization framework for FL in wireless IoT networks. Specifically, the framework supports client selection and wireless resource allocation, which includes the power and bandwidth allocation of the clients. In the framework, the objective function takes both the amount of training data and wireless resource consumption into consideration.
- We solve the formulated problem by using an alternative direction-based algorithm. In the algorithm, the primal problem is decomposed and transformed, and then, by combining and solving the constraints simultaneously, we derive the iteration

equation for the optimal power and bandwidth. After that, the optimal client selection can also be achieved using a greedy algorithm.

- We conduct extensive simulations using the MNIST to examine the effectiveness of the proposed joint optimization in utility of the network, accuracy rate and energy consumption.

The rest of the paper is organized as follows. Section 2 introduces the system model. Section 3 describes the problem formulation. The proposed algorithm is then presented in Section 4. The simulation results are discussed in Section 5. Finally, the conclusion is drawn in Section 6.

2. System Model

In this section, we describe the system model of FL-aided wireless IoT networks, which is divided into the network model, learning process and communication model.

2.1. Network Model

As depicted in Figure 1, we consider a FL-aided wireless IoT network, which consists of one base station (BS) and M IoT nodes. Let $\mathcal{M} = \{1, \cdots, M\}$ denote the set of IoT nodes. A FL server is directly deployed at BS, which is responsible for broadcasting and aggregating FL model parameters. In the FL-aided wireless IoT network, a global learning model of interest is trained by the cooperation of the FL server and IoT nodes. In the training, each IoT node acts as a FL client, which collects data as the training sample for its local training. Let $S_i = \{S_1, \cdots, S_l, \cdots, |S_i|\}$ denote the collected training data by IoT node i, where S_l is the lth training sample of the data. Thus, the whole training data set can be expressed as $S = \sum_{i \in \mathcal{M}} S_i$.

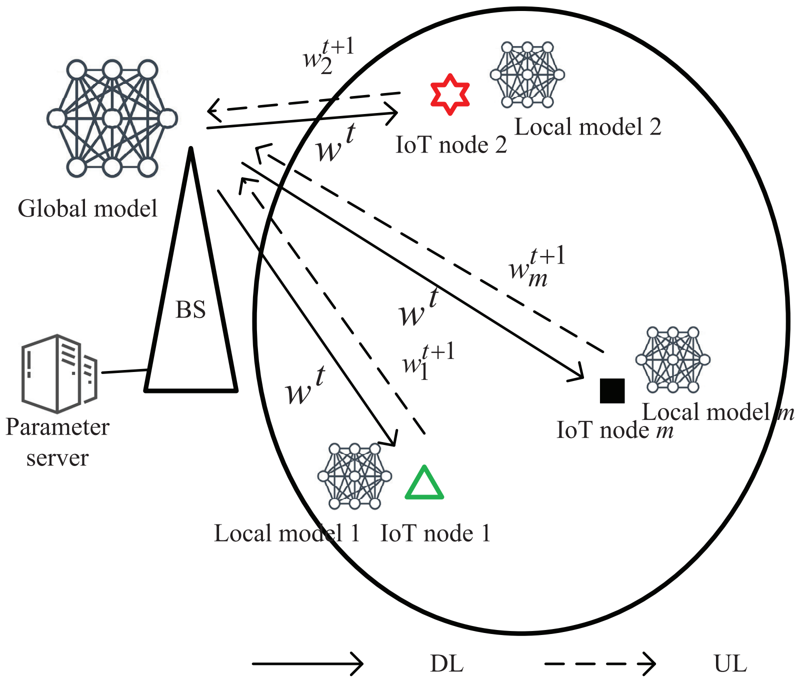

Figure 1. System model of FL-aided wireless IoT networks.

2.2. Learning Process

In the considered network, an ML model of interest is trained in a distributed manner by the cooperation of the FL server and IoT nodes. The goal of the training is to obtain the model parameter $\mathbf{w}$ by minimizing the loss function $f(\mathbf{w})$ on the data set S. The minimization can be expressed as

$$\min_{\mathbf{w}} f(\mathbf{w}) \triangleq \min_{\mathbf{w}} \frac{1}{|S|} \sum_{i=1}^{M} \sum_{l=1}^{|S_i|} f_i(\mathbf{w}, S_l), \tag{1}$$

where $f_i(\mathbf{w}, S_l)$ is the local loss function of IoT node i on sample S_l. We focus on the widely used federated averaging learning framework [2], where the training communication round is periodical. The tth communication round consists of the broadcasting phase, the local training phase and the aggregating phase. In the broadcasting phase, the FL server broadcasts the global model parameter $\mathbf{w}^t$ to all the IoT nodes via the wireless down link (DL) of BS. Then, in the local training phase, each IoT node i calculates the gradient of the local loss function $\nabla f_i(\mathbf{w}^t, S_i)$, and then, E epochs of the gradient descent method are used to obtain $\mathbf{w}^{t+1}$ as

$$\mathbf{w}_i^{t+1} = \mathbf{w}^t - \zeta_i \nabla f_i(\mathbf{w}^t, S_i), \tag{2}$$

where ζ_i is the learning rate of IoT node i. Finally, in the aggregating phase, each IoT node i transmits its local model parameter $\mathbf{w}_i^{t+1}$ to BS via the wireless up link (UL), while the FL server updates the global model parameter as

$$\mathbf{w}^{t+1} = \sum_{i=1}^{M} \frac{|S_i|}{|S|} \mathbf{w}_i^{t+1}. \tag{3}$$

The learning process terminates when the following condition holds.

$$f(\mathbf{w}^t) - f(\mathbf{w}^{t+1}) \le \Lambda, \tag{4}$$

where Λ is the learning termination threshold.

2.3. Communication Model

In the learning process, the aggregating phase is the bottleneck, because the wireless UL of the IoT node is resource-limited compared to the DL of BS. Hence, we focus on the resource optimization of the wireless UL in FL-aided wireless IoT networks. We assume that each IoT node transmits its local model parameter through frequency division multiple access. The achievable rate of IoT node i at the tth round can be written as

$$r_i^t = B_i^t \log_2(1 + \frac{P_i^t h_i^t}{\sigma^2}), \tag{5}$$

where B_i^t is the frequency bandwidth of IoT node i. P_i^t is the transmitting power of IoT i, h_i^t is the channel gain from IoT node i to BS, σ^2 is the power spectral density of the background noise. For each IoT node i, the local model parameter $\mathbf{w}_i^{t+1}$ is encapsulated into a packet with a size of D in the transmission. Thus, the transmission time from IoT node i to BS can be written as

$$T_i^t = \frac{D}{r_i^t}. \tag{6}$$

With (6), the outage probability of IoT node i can be expressed as

$$p_i^t = \Pr(T_i^t > \theta), \tag{7}$$

where θ is the maximum acceptable latency for the tth round in FL.

For the convenience of readers, the parameters concerning the formulation can be found in Table 1.

Table 1. Explanation of Abbreviations.

Notation	Definition
$\mathcal{M}$	The set of IoT nodes
B_i^t	The bandwidth of IoT node i in tth communication round
S_i	The collected data by IoT node i
S_l	The lth training sample of the data
P_i^t	The transmitting power of IoT node i in tth communication round
x_i^t	The node selection indicator of IoT node i in tth communication round
p_i^t	The outage probability of IoT node i in tth communication round
r_i^t	The achievable rate of IoT node i in the tth communication round

3. Problem Formulation

To improve the accuracy of the trained model, it is desirable to bring as much data as possible into the training process [19]. Meanwhile, it is also beneficial to reduce the consumption of wireless resources in parameter transmission. Hence, the utility function of the FL-aided wireless IoT network is formulated as

$$U_{\mathrm{FL}} = \sum_{i \in M} x_i^t b_i^t |S_i| - \tau \sum_{i \in M} x_i^t b_i^t B_i^t P_i^t, \tag{8}$$

where x_i^t is the node selection indicator, $x_i^t = 1$ represents that IoT node i is selected to participate in the tth round training; otherwise, $x_i^t = 0$. b_i^t is the outage indicator, $b_i^t = 1$ indicates that the UL of node i experiences an outage; otherwise, $b_i^t = 0$. τ denotes the cost coefficient. In (8), the first term captures the total amount of training data, while the second term represents the communication cost, which is formulated by the bandwidth-power product [20].

Our goal is to optimize the bandwidth, power and client selection by maximizing the utility function U_{FL}; therefore, the problem is formulated as follows.

$$\mathbf{P1:} \quad \max_{\mathbf{x},\mathbf{B},\mathbf{P}} \; U_{\mathrm{FL}}$$

$$\text{s.t.} \quad \text{C1:} \; x_i^t \in \{0,1\}, \;\; \forall i \in \mathcal{M}, \tag{9}$$

$$\text{C2:} \; \sum_{i \in \mathcal{M}} B_i^t x_i^t \le B_T, \tag{10}$$

$$\text{C3:} \; P_i^t \in [0, P_M], \;\; \forall i \in \mathcal{M}, \tag{11}$$

$$\text{C4:} \; b_i^t = \begin{cases} 1, & T_i^t \le \theta, i \in \mathcal{M}, \\ 0, & T_i^t > \theta, i \in \mathcal{M}. \end{cases} \tag{12}$$

Constraint C1 is the binary integer constraint for the client selection indicator. Constraint C2 guarantees that the sum of the allocated bandwidth cannot be beyond the total bandwidth. Constraint C3 ensures that the transmitting power of each IoT node is non-negative and cannot be beyond the maximal value. Constraint C4 denotes that only when the transmission time is smaller than the maximum acceptable latency can the transmission avoid the outage.

4. Algorithm Design

It can be observed that **P1** belongs to mixed integer non-linear programming (MINLP); therefore, its complexity is high. In order to handle this intractable problem, we propose an alternative direction-based algorithm. In the proposed algorithm, $\mathbf{x}$ is first fixed to **1** while **B** and **P** are solved from the simplified problem. After that, the obtained **B** and **P** are brought into **P1** to solve $\mathbf{x}$.

4.1. Solving **B** and **P** When **x** Is Given

When $\mathbf{x}$ is fixed to $\mathbf{1}$, it indicates that all the nodes are selected to join the learning process. In that case, **P1** can be simplified to

$$\begin{aligned} \textbf{P2:} \quad & \min_{\mathbf{B},\mathbf{P}} \; f_{\mathrm{P2}} = \sum_{i \in \mathcal{M}} B_i^t P_i^t \\ \text{s.t.} \quad & \sum_{i \in \mathcal{M}} B_i^t \leq B_T, \\ & \mathrm{C3}. \end{aligned} \tag{13}$$

Comparing constraint C4 with the objective function f_{P2}, it can be derived that $T_i^t = \theta$ should be satisfied, so by combining (5) and (6), there is

$$\frac{D}{\theta \log_2(1 + \frac{P_i^t h_i^t}{\sigma^2})} = B_i^t. \tag{14}$$

By substituting (14) into **P2**, there is

$$\begin{aligned} \textbf{P3:} \quad & \min_{\mathbf{P}} \; f_{\mathrm{P3}} = \sum_{i \in \mathcal{M}} \frac{D}{\theta \log_2(1 + \frac{P_i^t h_i^t}{\sigma^2})} P_i^t \\ \text{s.t.} \quad & \mathrm{C3}. \end{aligned} \tag{15}$$

Using the derivation $\frac{df_{\mathrm{P3}}}{dP_i} = 0$, it can be derived that

$$D\theta \log_2(1 + \frac{P_i^t h_i^t}{\sigma^2}) - \frac{DP_i\theta}{\ln 2} \frac{\frac{P_i^t h_i^t}{\sigma^2}}{1 + \frac{P_i^t h_i^t}{\sigma^2}} = 0. \tag{16}$$

Hence, we utilize the iteration method in [21] and derive the iteration equation of P_i as

$$P_i^{(k+1)t} = \frac{\log 2(1 + \frac{P_i^{(k)t} h_i^t}{\sigma^2}) \ln 2(\sigma^2 + P_i^{(k)t} h_i^t)}{P_i^{(k)t} h_i^t}, \tag{17}$$

where k is the iteration index. We denote P_i^* as the optimal value of P_i^t when the iteration (17) ends. By substituting P_i^* into (14), there is

$$B_i^* = \frac{D}{\theta \log_2(1 + \frac{P_i^* h_i^t}{\sigma^2})}, \tag{18}$$

where B_i^* denotes the optimal value of B_i^t.

4.2. Solving **x** with **B** and **P**

By substituting the obtained B_i^* and P_i^* into **P1**, it can be simplified as

$$\begin{aligned} \textbf{P4:} \quad & \max_{\mathbf{x}} \; \sum_{i \in \mathcal{M}} (|S_i| - \tau B_i^* P_i^*) x_i^t \\ \text{s.t.} \quad & \mathrm{C1} \sim \mathrm{C2}. \end{aligned} \tag{19}$$

Since **P4** belongs to the knapsack problem, we propose a greedy algorithm to solve it in Algorithm 1.

Algorithm 1 Greedy Algorithm for **P4**.

Input: $\mathcal{M}, B_T, |S_i|, B_i^*, P_i^*$
Initialize: $W_T = 0, x_i = 0, \Phi = \mathcal{M}$
for $i = 1 : M$ **do**
  if $W_T \leq B_T$ **then**
    $j = \arg\max_{i\in\Phi} \frac{(|S_i| - \tau B_i^* P_i^*)}{B_i^*}$
    $W_T = W_T + B_j^*$
    $x_j = 1$
    $\Phi = \Phi - \{j\}$
  end if
end for
Output: $\mathbf{x}$

With Algorithm 1, the proposed client and resource optimization can be summarized as Algorithm 2.

Algorithm 2 Client and Resource Optimization for FL-aided IoT.

Input: $\mathcal{M}, h_i^t, B_T, P_M, \theta$
Initialize: $\mathbf{x} = \mathbf{1}$
for $i = 1 : M$ **do**
  compute P_i^t by iteration (17)
  compute B_i^t by (18)
end for
run **Algorithm 1** to obtain x_i^t
Output: $\mathbf{x}, \mathbf{B}, \mathbf{P}$

The block diagram of the proposed algorithm is presented in Figure 2. We propose an alternative direction-based algorithm to handle the high complexity of **P1**. First, $\mathbf{x}$ is fixed to $\mathbf{1}$, and **P1** can be simplified to **P2**. Then by combining and substituting C4, (5) and (6) into **P2**, it can be further simplified to **P3**. Letting the derivation $\frac{df_{\mathrm{P3}}}{dP_i} = 0$ in **P3**, the iteration Equation (17) is derived. Through the iteration of (17), P_i^* can be obtained. Then, by substituting P_i^* into (14), B_i^* is also solved. Next, by substituting and into **P1**, it can be transformed to **P4**, which belongs to the knapsack problem. Hence, we adopt Algorithm 1 to solve $\mathbf{x}$ from **P4**. After that, $\mathbf{x}$, P_i^* and B_i^* are obtained and used for the client and resource optimization.

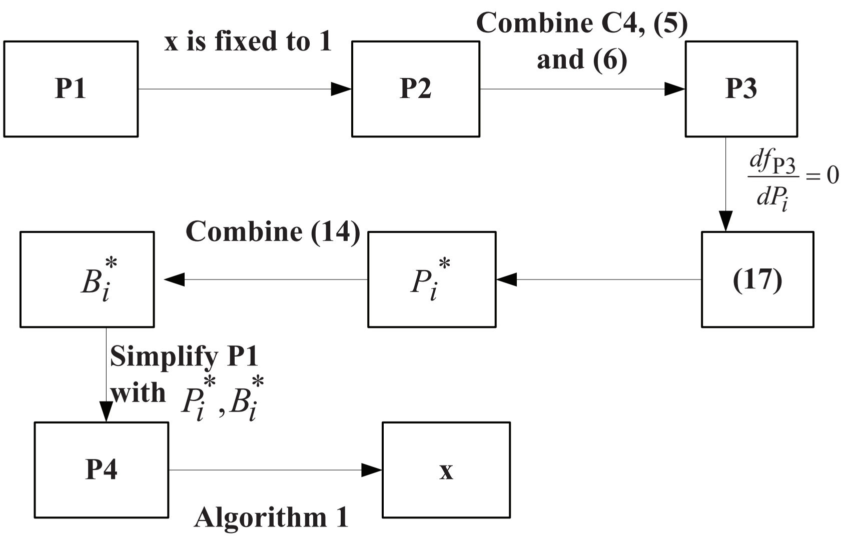

Figure 2. Block diagram of the proposed algorithm.

5. Simulation Results

In this section, the performance of the proposed alternative direction-based algorithm is examined by comparison with two existing benchmarks. Benchmark 1 is the vanilla FL scheme [2] over wireless IoT networks, among which the federated average algorithm is adopted. Benchmark 2 is the communication-efficient FL [18], among which the optimization objective is to maximize the number of clients participating in the FL process. We consider a FL-aided cellular IoT network within a 1 km $\times$ 1 km area, where the BS is located at the center while 20 IoT nodes are randomly distributed. The channel model [22] is adopted; that is, $103.2 + 27.3 \log_{10}(d)$ is adopted between BS and the IoT nodes. The total bandwidth of the network $B_T = 10$ MHz, and the power of background noise $\sigma^2 = -109$ dBm. The maximum transmitting power $P_M = 15$ dBm. In the simulation, the ML model of interest is a convolutional neural network (CNN) [23], which consists of 2 convolutional layers, 1 fully connected layer, and 1 softmax function-based output layer. Each convolutional layer is 5×5 and is connected with a 2×2 max-pooling. The fully connected layer has 500 units. The training data set is MNIST [24], and the original data from MNIST are randomly partitioned into M pieces and each IoT node is assigned one piece. Thus, the data distribution follows i. i. d. Here, the amount of training data S_i follows uniform distribution between 1500$\sim$3500 images.

Figure 3 shows the comparisons of network utility as the number of transmission round increases. It is observed that as the number of transmission round grows, the proposed algorithm outperforms benchmark 2. In fact, in the optimization of benchmark 2, the objective function considers all the nodes equally, while in the proposed algorithm, the amount of training data collected by the IoT nodes is considered as the profit weight in the objective function, which leads to more data being brought into the training under the same communication rounds. Meanwhile, the proposed algorithm and benchmark 2 are superior to benchmark 1. The reason is that in these two algorithms, the power and bandwidth are optimized to maintain the learning process, which efficiently reduces the consumption of communication resources.

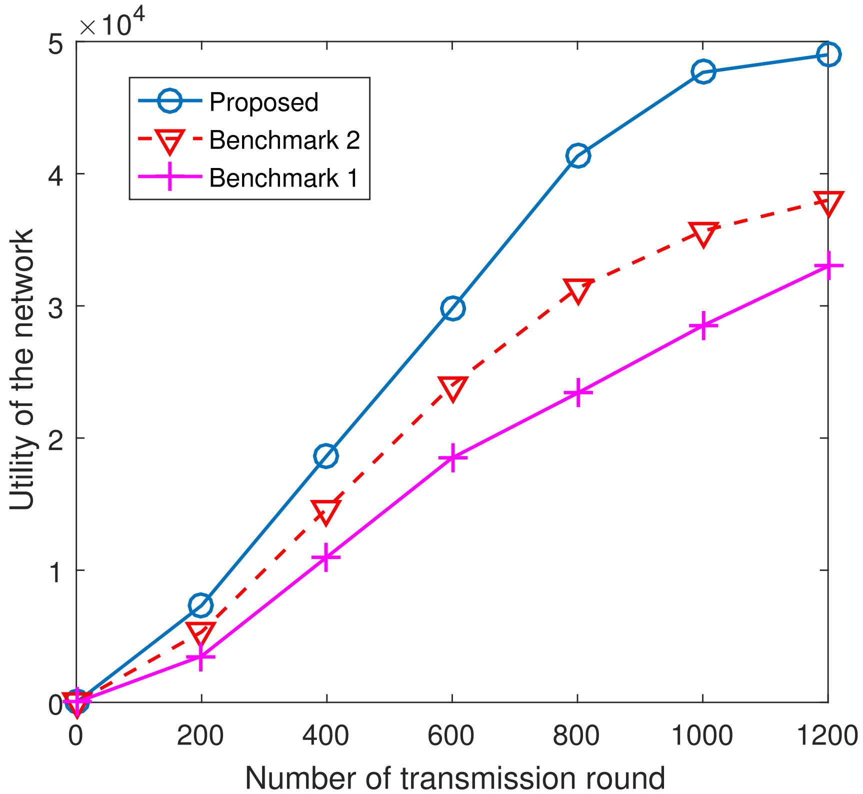

Figure 3. Comparisons of network utility when the number of transmission round increases.

Figure 4 depicts the loss of the trained model for the compared algorithms. First, as the number of transmission rounds increases, the loss of the trained model decreases for all three algorithms. The reason is that with the growth of the transmission round, more and more data are brought into training, which is helpful to improve the trained parameters. We found that when the number of transmission rounds is larger than 1000, the proposed algorithm and benchmark 2 converge to almost the same loss level. This is due to the fact that with sufficient communication rounds, the two schemes can bring adequate data into the training. However, recalling the results in Figure 3, it can be deduced that the consumption of communication resources are different for the two algorithms, and the proposed algorithm is superior in power and bandwidth saving.

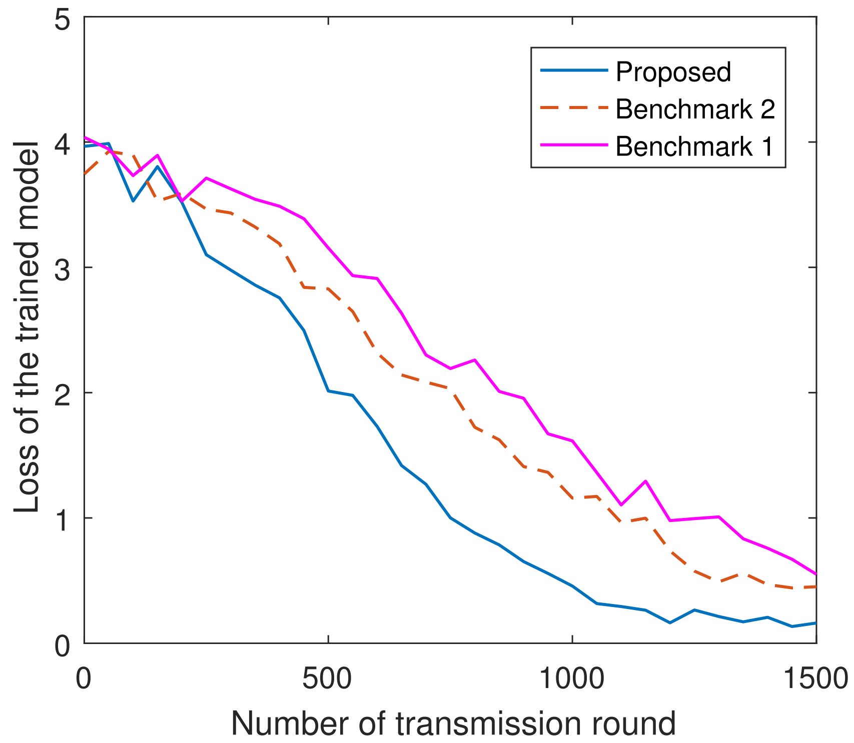

Figure 4. The loss of the trained model for the compared algorithms.

Figure 5 illustrates the comparisons of the accuracy rate when the number of transmission rounds increases. It is observed that as the transmission rounds grow, the accuracy rate of the trained model also increases for all compared algorithms. The proposed algorithm and benchmark 2 are superior to benchmark 1. The reason is that compared to benchmark 1, the two schemes brought more data into training with limited communication resources. When the number of transmission rounds is greater than 1000, the performance of the three algorithms is close, because in that case, sufficient data have been brought into the training process. The proposed algorithm outperforms benchmark 2 because its objective function considers both the amount of training data and the consumption of communication resources. As a result, compared to benchmark 2, the proposed algorithm efficiently brings more data into the training process.

Figure 6 depicts the comparisons of energy consumption per client when the number of transmission rounds increases. We observed that when the number of transmission rounds increases, the energy consumption per client also increases for all the three compared algorithms; therefore, the energy consumption is basically proportional to the number of transmissions. The proposed algorithm consumes less energy than the two benchmarks because the proposed optimization only selects the client with a superior wireless channel state and a large amount of local training data, which reduces the energy consumption in parameter transmission. On the contrary, benchmark 1 consumes more energy than benchmark 2 and the proposed algorithm because its client selection is random without considering wireless channel state. Additionally, benchmark 2 also consumes more energy than the proposed algorithm. The reason is that the optimization objective function of benchmark 2 has not taken the amount of local training data into consideration. This result indicates that the proposed algorithm achieves superior trade-off between the wireless channel state and the amount of local training data in parameter transmission.

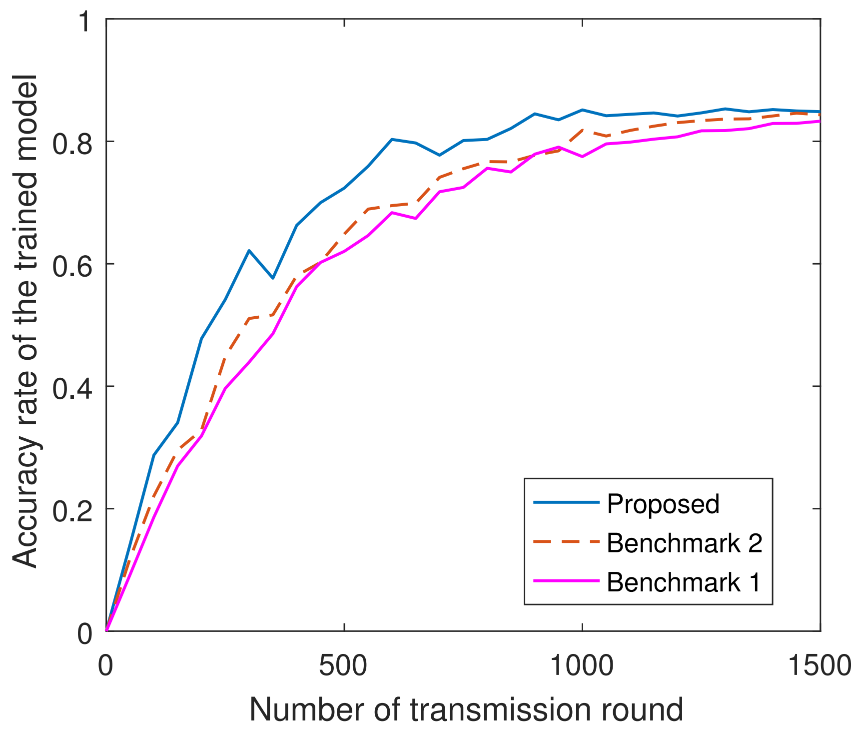

Figure 5. Comparisons of accuracy rate when the number of transmission rounds increases.

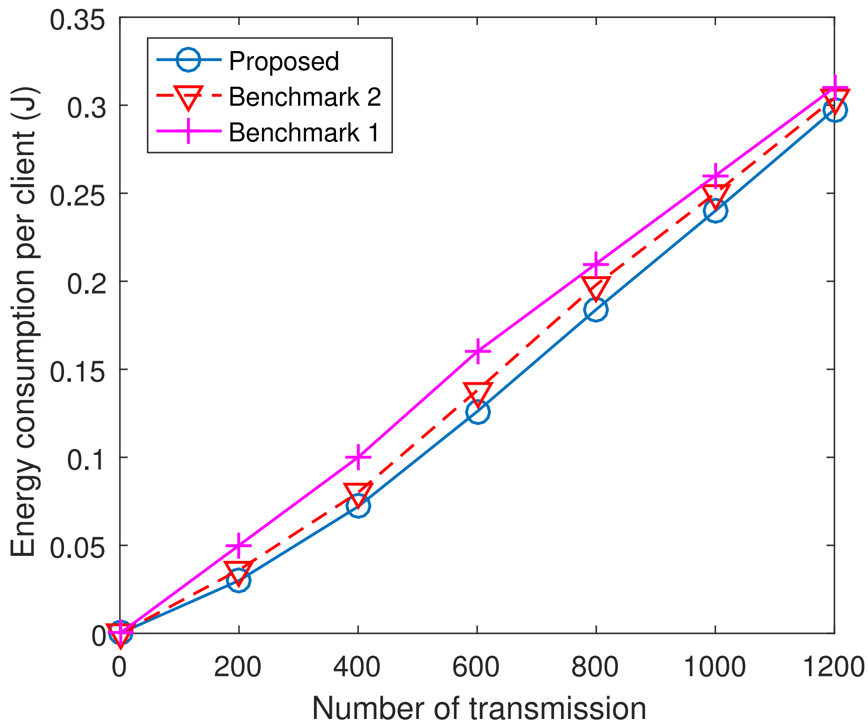

Figure 6. Comparisons of energy consumption per client when the number of transmission rounds increases.

6. Conclusions

In this paper, we have investigated the joint optimization of client selection and communication resource allocation in FL-aided wireless IoT networks. By taking both the amount of trained data and the consumption of communication resources into consideration, a MINLP was formulated to maximize the utility of the network. Then, an alternative direction-based algorithm was proposed to solve the problem efficiently. The simulation results have confirmed that the proposed algorithm is effective in reducing communication resource consumption and improving learning performance. Furthermore, the results also revealed that distinguishing node weights based on the amount of collected data is beneficial for FL. In the future work, we will extend the algorithm to the FL-enabled scenario with massive multiple-input multiple-output communications, where the higher dimensional signals make joint optimization of client selection and resource allocation more challenging.

Author Contributions: Conceptualization, J.Z. and Y.N.; methodology, J.Z. and Y.C.; validation, J.Z. and Y.N.; formal analysis, J.Z.; algorithm investigation, J.Z. and Y.C.; writing—original draft preparation, J.Z.; writing—review and editing, J.Z., Y.N. and Y.C.; supervision, Y.N.; project administration, Y.N.; funding acquisition, Y.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part by the Jiangsu University Philosophy and Social Science Research Fund under Grant 2022SJYB0517, in part by the Open Research Fund Project of Jiangsu Provincial Key Laboratory of Wireless Communication at Nanjing University of Posts and Telecommunications: Research on Key Technologies of Backscattering Communication Assisted by Intelligent Reflective Surface, and in part by the Natural Science Foundation on Frontier Leading Technology Basic Research Project of Jiangsu under Grant BK20212001.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy concerns.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Imteaj, A.; Thakker, U.; Wang, S.; Li, J.; Hadi, A.M. A Survey on Federated Learning for Resource-Constrained IoT Devices. *IEEE Internet Things J.* **2022**, *9*, 1–24. [CrossRef]
2. McMahan, H.B.; Moore, E.; Ramage, D.; Hampson, S.; Arcas, B.A.Y. Communication-efficient Learning of Deep Networks from Decentralized Data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, Ft. Lauderdale, FL, USA, 20–22 April 2017.
3. Zhou, Y.; Shi, Y.; Zhou, H.; Wang, J.; Fu, L.; Yang, Y. Toward Scalable Wireless Federated Learning: Challenges and Solutions. *IEEE Internet Things Mag.* **2023**, *6*, 10–16. [CrossRef]
4. Mishra, K.; Puthal, D. Data Sampling in Federated Learning: Principles, Features and Taxonomy. *IEEE Commun. Stand. Mag.* **2023**, *7*, 28–33. [CrossRef]
5. Xu, J.; Wang, H. Client Selection and Bandwidth Allocation in Wireless Federated Learning Networks: A Long-Term Perspective. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 1188–1200. [CrossRef]
6. Nishio, T.; Yonetani, R. Client Selection for Federated Learning with Heterogeneous Resources in Mobile Edge. In Proceedings of the 2019 IEEE International Conference on Communications, Shanghai, China, 20–24 May 2019.
7. Amiri, M.M.; Gündüz, D.; Kulkarni, S.R.; Poor, H.V. Update Aware Device Scheduling for Federated Learning at the Wireless Edge. In Proceedings of the 2020 IEEE International Symposium on Information Theory, Los Angeles, CA, USA, 21–26 June 2020.
8. Chen, M.; Yang, Z.; Saad, W.; Yin, C.; Poor, H.V.; Cui, S. A Joint Learning and Communications Framework for Federated Learning over Wireless Networks. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 269–283. [CrossRef]
9. Liu, X.; Deng, Y.; Mahmoodi, T. Wireless Distributed Learning: A New Hybrid Split and Federated Learning Approach. *IEEE Trans. Wirel. Commun.* **2023**, *22*, 2650–2665. [CrossRef]
10. Wang, P.; Sun, W.; Zhang, H.; Ma, W.; Zhang, Y. Distributed and Secure Federated Learning for Wireless Computing Power Networks. *IEEE Trans. Veh. Technol.* **2023**, *72*, 9381–9393. [CrossRef]
11. Sun, W.; Li, Z.; Wang, Q.; Zhang, Y. FedTAR: Task and Resource-Aware Federated Learning for Wireless Computing Power Networks. *IEEE Internet Things J.* **2023**, *10*, 4257–4270. [CrossRef]
12. Yao, J.; Yang, Z.; Xu, W.; Chen, M.; Niyato, D. GoMORE: Global Model Reuse for Resource-Constrained Wireless Federated Learning. *IEEE Wirel. Commun. Lett.* **2023**, *12*, 1543–1547. [CrossRef]

13. Zhu, H.; Zhou, Y.; Qian, H.; Shi, Y.; Chen, X.; Yang, Y. Online Client Selection for Asynchronous Federated Learning with Fairness Consideration. *IEEE Trans. Wirel. Commun.* **2023**, *22*, 2493–2506. [CrossRef]
14. Hu, C.; Chen, Z.; Larsson, E.G. Scheduling and Aggregation Design for Asynchronous Federated Learning over Wireless Networks. *IEEE J. Sel. Areas Commun.* **2023**, *41*, 874–886. [CrossRef]
15. You, C.; Guo, K.; Yang H.; Quek, T.Q.S. Hierarchical Personalized Federated Learning over Massive Mobile Edge Computing Networks. *IEEE Trans. Wirel. Commun.* **2023**, *22*, 8141–8157. [CrossRef]
16. Lee, H. Device Selection and Resource Allocation for Layerwise Federated Learning in Wireless Networks. *IEEE Syst. J.* **2022**, *16*, 6441–6444. [CrossRef]
17. Wen, W.; Chen, Z.; Yang, H.; Xia, W.; Quek, T.Q.S. Joint Scheduling and Resource Allocation for Hierarchical Federated Edge Learning. *IEEE Trans. Wirel. Commun.* **2022**, *21*, 5857–5872. [CrossRef]
18. Chen, H.; Huang, S.; Zhang, D.; Xiao, M.; Skoglund, M.; Poor, H.V. Federated Learning over Wireless IoT Networks with Optimized Communication and Resources. *IEEE Internet Things J.* **2022**, *9*, 16592–16605. [CrossRef]
19. Chen, Y.; Sun, X.; Jin, Y. Communication-Efficient Federated Deep Learning With Layerwise Asynchronous Model Update and Temporally Weighted Aggregation. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *31*, 4229–4238. [CrossRef] [PubMed]
20. Li, H.; Wang, K. Weighted Bandwidth–Power Product Optimization in Downlink Femtocell Networks. *IEEE Commun. Lett.* **2015**, *19*, 1588–1591. [CrossRef]
21. Yao, J.; Ansari, N. Caching in Energy Harvesting Aided Internet of Things: A Game-Theoretic Approach. *IEEE Internet Things J.* **2019**, *6*, 3194—3201. [CrossRef]
22. Liu, X.; Ansari, N. Dual-Battery Enabled Profit Driven User Association in Green Heterogeneous Cellular Networks. *IEEE Trans. Green Commun. Netw.* **2018**, *20*, 1002–1011. [CrossRef]
23. Poudyal, A.; Tamrakar, U.; Trevizan, R.D.; Fourney, R.; Tonkoski, R.; Hansen, T.M. Multiarea Inertia Estimation Using Convolutional Neural Networks and Federated Learning. *IEEE Syst. J.* **2022**, *16*, 6401–6412. [CrossRef]
24. Deng, L. The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]. *IEEE Signal Process Mag.* **2012**, *29*, 141–142. [CrossRef]