IEEE Transactions on Cognitive Communications and Networking

Data-Aware Device Scheduling for Federated Edge Learning

Presented by:
Ittehad Saleh Chowdhury(11)
Reyadath Ullah(33)

Authors:
Afaf Taik
Zoubeir Mlika,
Soumaya Cherkaoui

Brief Recap

Previous Reviewed Paper: "***Energy Efficient User Scheduling for Hybrid Split and Federated Learning in Wireless UAV Networks***"

The main objective and contribution of the paper was:

- Energy + Resource Optimization
- User Scheduling
- Learning Under Unreliable Wireless Channels

Why this paper?

The previous paper used a term denoted as the “***Diversity Index***”. The claim was that, “***Diversity Index***” provides a method for quantifying the diversity of users within a wireless network but very little context was actually provided.

We hope to get a better understanding of the formulation of this index and how it affects the learning speed and eff-iciency.

$$
\begin{aligned}
\mathcal{OP}_1 \quad & \min_{b_i,\,\{\alpha_{ij}\}} \sum_{i=1}^{N} \sum_{j \in \mathcal{J}} \alpha_{ij}(E_{ij} - I_{ij}) \\
s.t. \quad & C_1 : \sum_{j \in \mathcal{J}} \alpha_{ij}\tau_{ij} \leq T, \\
& C_2 : \alpha_{ij} \in \{0,1\},\ \forall j \in \mathcal{J},\ \forall i \in \mathcal{N}, \\
& C_3 : \sum_{j \in \mathcal{J}} \alpha_{ij} \leq 1,\ \forall i \in \mathcal{N}, \\
& C_4 : \sum_{i=1}^{N} \sum_{j \in \mathcal{J}} \alpha_{ij} b_i \leq 1, \\
& C_5 : 0 \leq b_i \leq 1,\ \forall i \in \mathcal{N},
\end{aligned}
$$

Introduction

The paper discusses Federated Edge Learning (FEEL) as a means to collaboratively train machine learning models across edge devices.

Challenges Highlighted:

- Limited communication bandwidth.
- Diverse data distribution across devices.

Solution Proposed:

A new scheduling algorithm that prioritizes devices based on the diversity of their data.

Primary Contributions

- Designing a suitable diversity indicator

- Formulate a joint device selection and bandwidth allocation problem taking with the baseline criterion being the data diversity

- Prove that the formulated problem is NP-Hard and propose a "***Data-Aware***" scheduling algorithm.

Diversity in Federated Learning

- Diversity on the construction of training batches improve the efficiency of the learning process.
- Additionally, "Active Learning" demonstrates how models can be trained with less data points given the selected samples are diverse and more informative.

We investigate the possibility of exploiting the different dataset properties to carefully select devices with potentially more informative datasets with less redundancy, by measuring their size and diversity.

Measures for Data Diversity

Simpson Index

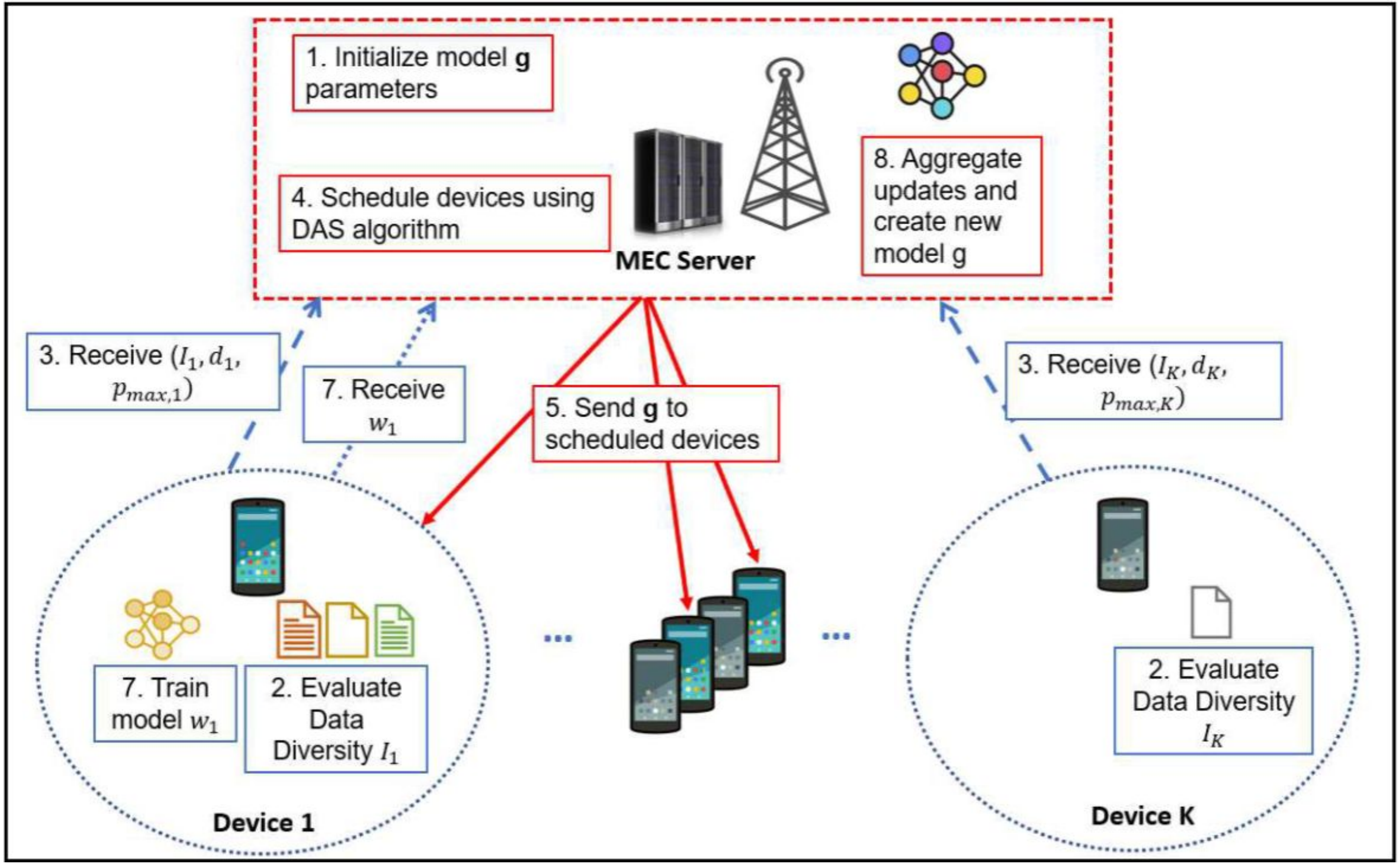
$$\lambda = \sum_{c=1}^{C} p_c^2,$$

The Gini–Simpson index is its transformation **1 − λ** , which represents the probability that the two samples belong to different classes.

Shannon Entropy

$$H = -\sum_{c=1}^{C} p_c \log_2(p_c),$$

System Model and Problem Formulation



1. Initialize model **g** parameters
4. Schedule devices using DAS algorithm
MEC Server
8. Aggregate updates and create new model g
3. Receive $(I_1, d_1, p_{max,1})$
7. Receive w_1
5. Send **g** to scheduled devices
3. Receive $(I_K, d_K, p_{max,K})$
...
...
7. Train model w_1
2. Evaluate Data Diversity I_1
Device 1
2. Evaluate Data Diversity I_K
Device K

Learning Model

Algorithm 1 FEEL Procedure

while $r < r_{max}$ or $accuracy <$ desired accuracy **do**
    if $r = 0$ **then**
        initialize the model's parameters at the MEC server
    end if
    Receive devices information (transmit power, available data size, dataset diversity index)
    Schedule a subset S_r of devices with at least N devices using Algorithm 2
    for device $k \in S_r$ **do**
        k receives model g
        k trains on local data D_k for E epochs
        k sends updated model w_k to MEC server
    end for
    MEC server computes new global model using weighted average: $g \leftarrow \sum_{k \in S_r} \frac{D_k}{D_r} w_k$
    start next round $r \leftarrow r + 1$
end while

- MEC server initializes the model and selects devices based on data diversity and communication needs.
- Ensures a minimum number K of participants for sufficient data collection each round.
- Devices train locally using SGD and upload updated parameters to the MEC.
- MEC aggregates via “**FedAvg”,** iterates training across rounds until the desired accuracy or $\mathbf{r_{max}}$ has been reached.

Dataset Diversity Index Design

Goals of device selection:

1. Select devices with potentially informative datasets by evaluating size and diversity
2. Guarantee that the selected devices are diversified by adding 'age-of-update' to the diversity index

Formulation:

A function is used to compute a weighted rank value for each device, bounded between [**0, γ_i**]

where, $i \in$ **{dataset diversity, dataset size, age}.** The value of this function is given as follows:

- $\nu_i \times \gamma_i$ where γ_i is the adjustable weight for each metric assigned by the server and ν_i is the normalized value of the metric **i**

Dataset Diversity Index Design

$\boldsymbol{v}_\mathbf{i}$ is calculated as follows:

$$v_i = \frac{\text{measured value of metric } i}{\text{maximum for metric } i}.$$

We define the diversity index of dataset ***k*** as:

$$I_k = \sum_i v_{i,k} \gamma_{i,k},$$

We formulate the first goal of the device selection problem as:

$$\max_x \sum_{k=1}^{K} I_k x_k, \tag{5}$$

Transmission Model

We define $\boldsymbol{\alpha}$ **= [α_1, ..., α_k]**, where for each device **k**, α_k **∈ [0, 1]** is the bandwidth allocation ratio.

- The achievable rate of device **k** when transmitting to the BS is given by:

$$r_k = \alpha_k B \log_2(1 + \frac{g_k P_k}{\alpha_k B N_0}), \qquad \forall k \in [1, K],$$

- The round duration is given by:

$$T = \max((t_k^{train} + t_k^{up}) x_k), \qquad \forall k \in [1, K],$$

- The wireless transmit energy is given by:

$$E_k = P_k t_k^{up}.$$

Device Scheduling Problem Formulation

We define the following goals for the device scheduling algorithm:

- From the perspective of devices, it is desirable to consume the least amount of energy to carry the training and uploading tasks

$$\min_{x,\alpha} \sum_{k=1}^{K} x_k E_k. \tag{11}$$

- To avoid stragglers' problem, it is desirable for the MEC server to have short round duration. Thus, the minimization of the communication round.

$$\min_{x,\alpha} T. \tag{12}$$

- For accelerated learning we adopt the goal defined in Eq 5.

Device Scheduling Problem Formulation(Cont.)

By combining these three goals, the problem is formulated as a multi-objective optimization problem as follows:

$$\underset{x,\,\alpha}{\text{minimize}} \quad \left\{ \sum_{k=1}^{K} x_k E_k, T - \sum_{k=1}^{K} x_k I_k \right\} \tag{13a}$$

subject to

$$(t_k^{train} + t_k^{up}) x_k \le T, \qquad \forall k \in [1, K], \tag{13b}$$

$$\sum_{k=1}^{K} \alpha_k \le 1, \qquad \forall k \in [1, K], \tag{13c}$$

$$0 \le \alpha_k \le 1, \qquad \forall k \in [1, K], \tag{13d}$$

$$\sum_{k=1}^{K} x_k \ge N, \qquad \forall k \in [1, K], \tag{13e}$$

$$x_k \in \{0, 1\}, \qquad \forall k \in [1, K]. \tag{13f}$$

Scheduling Algorithm

Besides being NP-hard, problem (13) is a mixed integer nonlinear multi-objective problem that is hard to solve.

We try to solve this problem in two main steps:

1. First, problem (13) is decomposed into two subproblems
2. The proposed algorithm optimizes iteratively both subproblems.

Sub-problem 1

Sub-problem 1 is a selection problem in which we select the devices in order to optimize a weighted linear combination of the different objectives:

$$\underset{x}{\text{minimize}} \quad \lambda_E \sum_{k=1}^{K} x_k E_k + \quad \lambda_T T \quad - \lambda_I \sum_{k=1}^{K} x_k I_k \tag{14a}$$

subject to

$$x_k \in \{0,1\} \qquad \forall k \in [1,K] \qquad , \tag{14b}$$

$$\sum_{k=1}^{K} x_k \geq K \qquad \forall k \in [1,K] \tag{14c}$$

Where λ_E , λ_T, and λ_I are positive scaling constants used to:

1. Scale the value of the objective function
2. Combine the different conflicting objectives into a linear single one

Sub-problem 1 (Cont.)

To solve Sub-problem 1, we use relaxation and rounding. We relax the integer constraint x_k which convert the problem into a continuous optimization problem, which is generally easier to solve.

$$\underset{x}{\text{minimize}} \quad \lambda_E \sum_{k=1}^{K} x_k E_k + \quad \lambda_T T \quad - \lambda_I \sum_{k=1}^{K} x_k I_k \tag{16a}$$

subject to

$$0 \leq x_k \leq 1 \qquad \forall k \in [1, K] \tag{16b}$$

The continuous value of x_k can be viewed as the selection priority of the device k, therefore, if the condition (14c) is not satisfied, we set $x_k = 1$ for the N devices with highest priorities.

Sub-problem 2

- Device selection decision is fixed through solving Sub-problem 1
- The main objective of Sub-problem 2 is Bandwidth Allocation

We use a Weighting Factor ρ for the consumed energy and round's completion time.

$$\underset{\alpha}{\text{minimize}} \quad \rho \sum_{k=1}^{K} x_k E_k + \quad (1-\rho)T \tag{15a}$$

$$\text{subject to} \qquad \sum_{k=1}^{K} \alpha_k \leq 1, \qquad \forall k \in [1, K], \tag{15b}$$

$$0 \leq \alpha_k \leq 1, \qquad \forall k \in [1, K] \tag{15c}$$

This problem can be solved using various "off-the-shelf" solvers

Algorithm 2

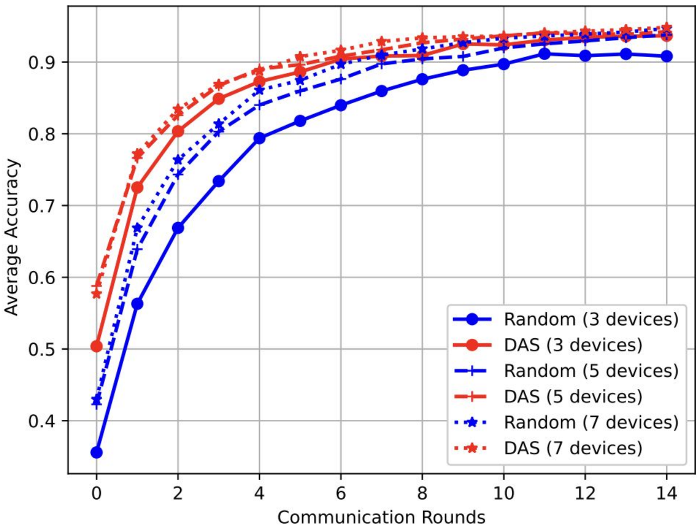
The proposed FEEL algorithm is an iterative algorithm that solves each sub-problem sequentially.

Algorithm 2 DAS algorithm for FEEL

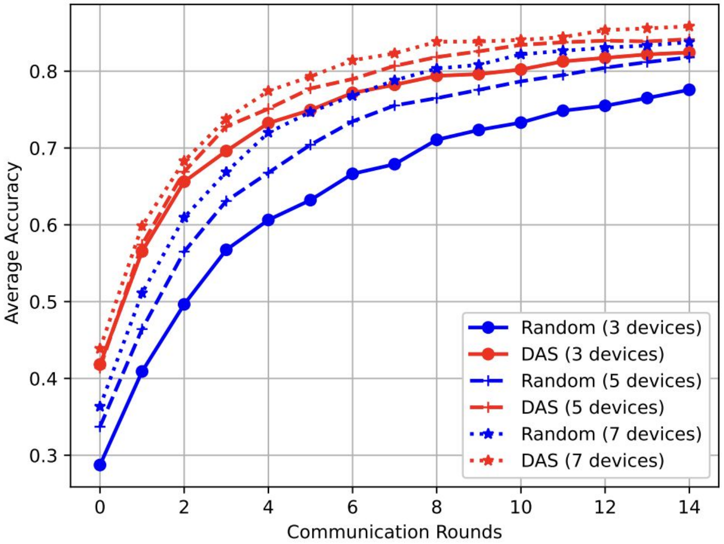
initialize $x_k = 1 \forall k \in [1, K]$
uniformly allocate the bandwidth
$iterations \leftarrow 1$
while $iterations < iterations_{max}$ **and not** convergence **do**
    Solve Sub1 : Return x
    Round x
    if condition(14c) is satisfied **then**
        continue
        select K devices with the highest priorities
    end if
    Solve Sub2 : Return α
    $iterations \leftarrow iterations + 1$
end while

Simulation and Results

Average test accuracy



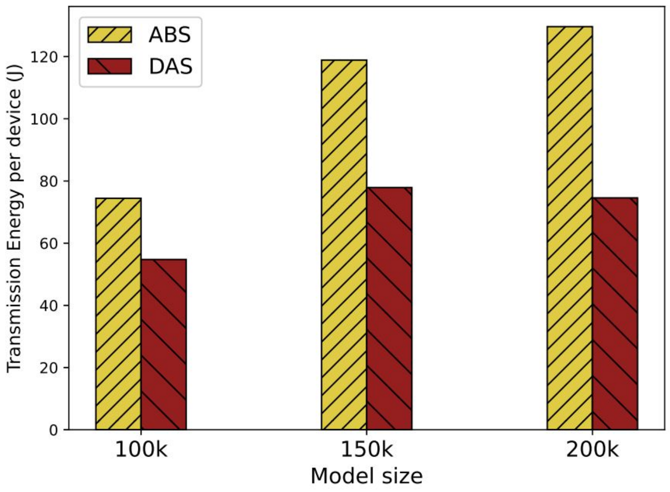
(a) CNN



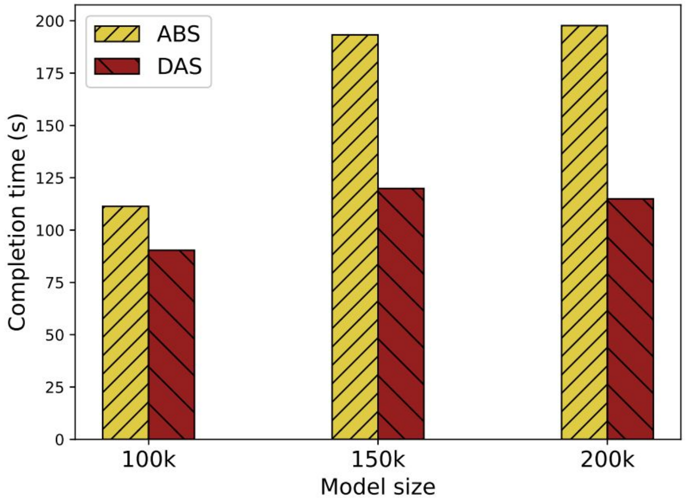
(b) MLP

Simulation and Results(Cont.)

Energy per device and completion time for training the CNN model for a goal accuracy of 92%



(a) Energy per device



(b) Completion time

Implications

$$
\begin{aligned}
\mathcal{OP}_1 \quad & \min_{b_i,\, \{\alpha_{ij}\}} \sum_{i=1}^{N} \sum_{j \in \mathcal{J}} \alpha_{ij}(E_{ij} - I_{ij}) \\
s.t. \quad & C_1 : \sum_{j \in \mathcal{J}} \alpha_{ij}\tau_{ij} \le T, \\
& C_2 : \alpha_{ij} \in \{0,1\},\ \forall j \in \mathcal{J},\ \forall i \in \mathcal{N}, \\
& C_3 : \sum_{j \in \mathcal{J}} \alpha_{ij} \le 1,\ \forall i \in \mathcal{N}, \\
& C_4 : \sum_{i=1}^{N} \sum_{j \in \mathcal{J}} \alpha_{ij} b_i \le 1, \\
& C_5 : 0 \le b_i \le 1,\ \forall i \in \mathcal{N},
\end{aligned}
\tag{6}
$$

Local Dataset Diversity:

- Age-of-Update
- Computation Capacity
- Local Dataset Size

Objective of this problem:

- Minimize Energy Consumption
- Maximize Data Diversity to increase the learning rate

Thank You!

Discussion and Observations