

EE441: Homework #3

Zhiyun Yu
zhiyunyu@buffalo.edu

University at Buffalo — February 11, 2025

1 Mac-Layer Aspects of Federated Learning (50 points)

Take a look at the paper titled *Federated Learning Over Wireless Networks: Convergence Analysis and Resource Allocation* posted along with lectures and answer the following questions:

Question 1

What are the main assumptions they are making regarding the loss functions? How realistic are those assumptions in machine learning applications?
(Hint: check out the text After Equation (1).

Ans: The paper makes the main assumption that the loss functions are strongly convex and smooth. This assumption is quite realistic in many machine learning applications because it ensures that the optimization problem has a unique global minimum and smooth gradients, which are essential for efficient convergence.

Question 2

How do they model the communication energy consumption and delay?
(Hint: check out section V-A). Note that their data-rate expression is using $\ln(\cdot)$ function, which as mentioned in the class leads to the unit of nats/sec for datarate. You can simply replace $\ln(\cdot)$ with $\log_2(\cdot)$ and datarate would become bits/sec, which is easier to understand.

Ans: Authors are model the communication energy consumption and delay by considering the data-rate expression using the natural logarithm (\ln) in Equation 14:

$$r_n = B \ln\left(1 + \frac{\bar{h}_n p_n}{N_0}\right),$$

where B is the bandwidth, N_0 is the background noise, p_n is the transmission power, and \bar{h}_n is the average channel gain of the UE n during the training time of Federal Learning.

Question 3

In their main problem formulation, what is the meaning behind the objective function? What tradeoff is it capturing?
(Hint: check out section V-B)

Ans: In Section V-B, authors are talking about the objective function of the formulation captures **the trade-off between the convergence wall-clock time and its energy consumption** of user equipments (i.e. In the article, its labeled as UE) with heterogeneous computing and power resources.

Question 4

What is the meaning behind the constraints imposed via Equations (18) and (19)?
(Hint: Think about what aspects of the system are they covering)

Ans:

Next to the first paragraph of V-B, authors seem to stated Equations 18

$$\sum_{n=1}^N \tau_n \leq T_{co}$$

and 19

$$\max_n \frac{c_n D_n}{f_n} = T_{cp}$$

impose constraints related to the communication energy consumption and delay in the federated learning system. While constraint (18) $\sum_{n=1}^N \tau_n \leq T_{co}$ captures the time-sharing uplink transmission of UEs, constraint (19) $\max_n \frac{c_n D_n}{f_n} = T_{cp}$ defines that the computing time in one local round is determined by the “bottleneck” UE (e.g., with large datasize and low CPU frequency.)

Question 5

Assume that you want to write a paper that extends this highly cited journal paper and make its system model more realistic. You think about a few directions and come up with this idea: “I want to optimize the tradeoff between the energy consumption, delay, and model performance”. So, you take another look at the objective function of their optimization problem and see that although energy consumption and delay are captured, model performance is not explicitly there. What would you do to change the objective function of their optimization problem so you can achieve your goal?
(Hint: See Equation (6), which is the goal of the machine learning model and think about parameter ε . See that ε also appears in Equation (12) that will determine the value of K_g in their objective function. However, they are not considering ε to be a controllable/optimizable value. So, you have to somehow add it to the objective function and also add it to the optimization variables).

Ans: For extends this highly cited journal paper and make its system model more realistic first thing pop-up to my mind is optimize the tradeoff between the energy consumption, delay, and model performance, I will make the system model more realistic and achieve my goal of optimizing the tradeoff between energy consumption, I think I can start with the delay, model performance, I would modify the objective function to explicitly include the model performance parameter, ε ensures the tradeoff is captured and put it into the constraints and it will optimized other variables.

2 Physical-Layer Aspects of Federated Learning (50 points)

Take a look at the paper titled *A Joint Learning and Communications Framework for Federated Learning Over Wireless Networks* posted along with the lectures and answer the following questions:

Question 6

Take a look at Equation (3). How data rate is connected/tied to the notion of resource block (RB) offered via OFDMA?

Ans: Equation (3):

$$c_i^U(r_i, P_i) = \sum_{n=1}^R r_{i,n} B^U \mathbb{E}_{h_i} \left(\log_2 \left(1 + \frac{P_i h_i}{I_n + B^U N_0} \right) \right)$$

represents the data rate considering the resource block allocation and channel conditions in an OFDMA system where $r_i = [r_{i,1}, \dots, r_{i,R}]$ is an RB allocation vector with R being the total number of RBs. The uplink rate of user i transmitting its local FL model parameters to the BS is given by the equation 3.

Question 7

Check Equation (7) and the text around it. How can we decrease the probability of error given the controllable parameters of the wireless devices?

(Hint: One of the main controllable parameters is the transmit power)

Ans: In the considered system whenever the received local FL model contains errors, the BS will not use it for the update of the global model. Author assume that in some scenario the BS will not ask the corresponding users to resend their local FL models when the local models contain the data errors. Later, authors update this assumption in to equations 8, the global FL model can be written as:

$$\mathbf{g}(\mathbf{a}, \mathbf{P}, \mathbf{R}) = \frac{\sum_{i=1}^U K_i a_i w_i C(w_i)}{\sum_{i=1}^U K_i a_i C(w_i)}$$

Question 8

Check Equation (8) and the text around it.

What differences do you see between that equation and conventional aggregation rule of federated learning, which is given by Equation (2)?

Why such differences exist (i.e., why are they introduced)?

Ans: Equation (2):

$$\mathbf{g}_t = \frac{\sum_{i=1}^U K_i \mathbf{w}_i, t}{K}$$

Equation (8):

$$\mathbf{g}(\mathbf{a}, \mathbf{P}, \mathbf{R}) = \frac{\sum_{i=1}^U K_i a_i w_i C(w_i)}{\sum_{i=1}^U K_i a_i C(w_i)}$$

Equation (8) in the paper modifies the conventional aggregation rule of federated learning, represented by Equation (2), by introducing additional weighting factors $C(w_i)$ and a dual summation. Thru this modify to Equations 8, it will optimize the trade-off between convergence speed and resource usage.

Question 9

Taking a look at their main optimization problem, what is the objective function capturing (i.e., what is the goal of the optimization problem)?

Why do you think constraints in (11b) and (11e), which are the main contributions/nuances of this formulation, are introduced?

Ans: Their main optimization problem is minimizing the energy consumption and delay while ensuring the model accuracy in FL over Wireless Networks. This minimization problem includes optimizing transmit power allocation as well as resource allocation for each user. Constraint in **11b** indicates that each user can occupy only one RB for uplink data transmission, ensures that the sum of the binary variables over all n from 1 to r equals the binary variable a_i for each user in set Number of users, and **11e** mainly introduced that each uplink RB can be allocated to at most one user. These constraints are introduced

Question 10

Assume that you want to write a paper that extends this highly cited journal paper. You again come up with this idea: "I want to optimize the tradeoff between the energy consumption, delay, and model performance". You look back again at the objective function and see that only the model performance is captured in it. So, you would like to add the delay and energy consumption to it as well. How do you think that can be done?

(Hint: Take a look at constraints (11c) and (11d). In those constraints, the right hand sides, which dictate the energy consumption and delay of the algorithm, are assumed to be given/fixed. So, you need to make them controllable/optimizable and somehow include them in the objective function).

Ans: I may start with Constraints 11c and 11d in the original paper, Transform these constraints into optimization variables, may improvement the energy consumption and delay in terms of other system parameters, that means in it may optimize the delay and its consumptions.