USC University of Southern California
Information Sciences Institute

Towards Multi-Objective Statistically Fair Federated Learning

Ninareh Mehrabi, Cyprien de Lichy, John McKay, Cynthia He, William Campbell

alexa ai
NATURAL UNDERSTANDING

Motivation

- In Federated Learning (FL), where data is local to the clients and access to the sensitive attributes is a challenge, can we satisfy different statistical fairness metrics, audit, and verify clients' models?
- Can we satisfy multiple objectives including statistical fairness metrics in FL?
- Can we identify and mitigate the effect of uncooperative or adversarial clients who might inject malicious, unfair, and poor-quality models into the federated system and instead reward better clients?

Methodology

- To answer the above, server can use a validation set. This validation or verification step has a couple of advantages:
  1. It gives the server a dataset on which it can compute fairness measures with existing sensitive attributes.
  2. Server can compute scores for each client model and weight each client accordingly
  3. Validation set can audit the FL model with regards to any sensitive attribute.

Objective:

$$\min_w f(w) = \sum_{k=1}^{K} p_k F_k(w) \quad \text{where} \quad F_k(w) = \frac{1}{n_k} \sum_{ik=1}^{n_k} f_{ik}(w)$$

$$p_k = \frac{\sum_{j=1}^{J} \gamma_j s_{jk}}{\Gamma S}$$

FedVal Algorithm

Algorithm 1: FedVal Algorithm

Input: k number of clients; γ_j weight for each objective j; B local minibatch size; E number of local epochs; η learning rate.

Output: w final federated model.

Server Side:

initialize w_0

for t= *1,2,...* **do**

  for *each client k in parallel* **do**

    $w_{t+1}^k \leftarrow$ ClientUpdate(k,w_t)

    Validate client k's model w_{t+1}^k when temporarily aggregated with the global FL model w_t and calculate $\sum_{j=1}^{J} \gamma_j s_{jk}$.

  end

  Rank each client k based on their scores $\sum_{j=1}^{J} \gamma_j s_{jk}$ and assign the rank score rs_k to each client k. (//Optional step refer to the Ranking Algorithm for more details).

  $w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{rs_k}{\sum_{k=1}^{K} rs_k} w_{t+1}^k$ (//rs_k can be replaced by $\sum_{j=1}^{J} \gamma_j s_{jk}$ if the optional step is skipped.)

end

return w_{t+1}

Client Side:

ClientUpdate(k,w):

  for *each local epoch i from 1 to E* **do**

    for *each batch b with size B* **do**

      $w \leftarrow w - \eta \nabla \ell(w; b)$

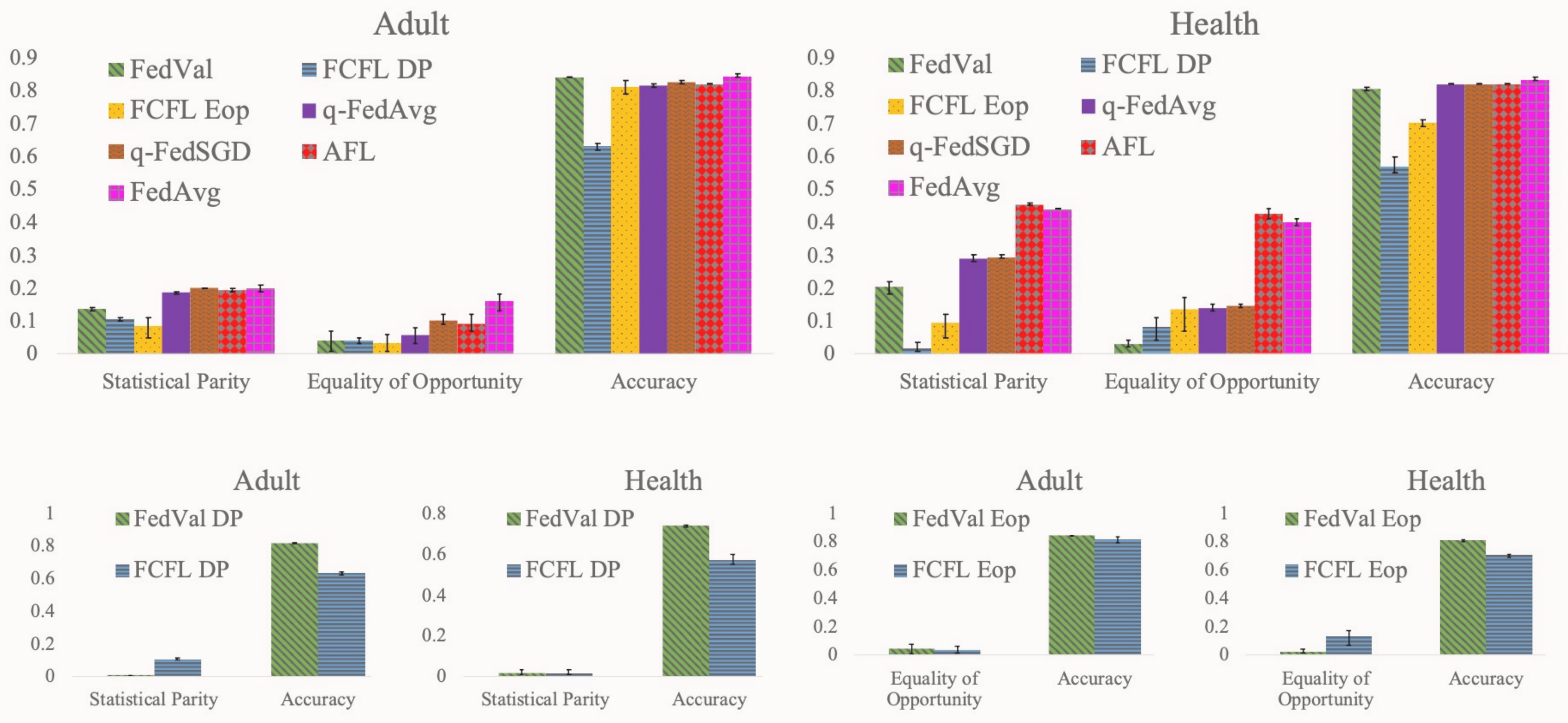
    end

  end

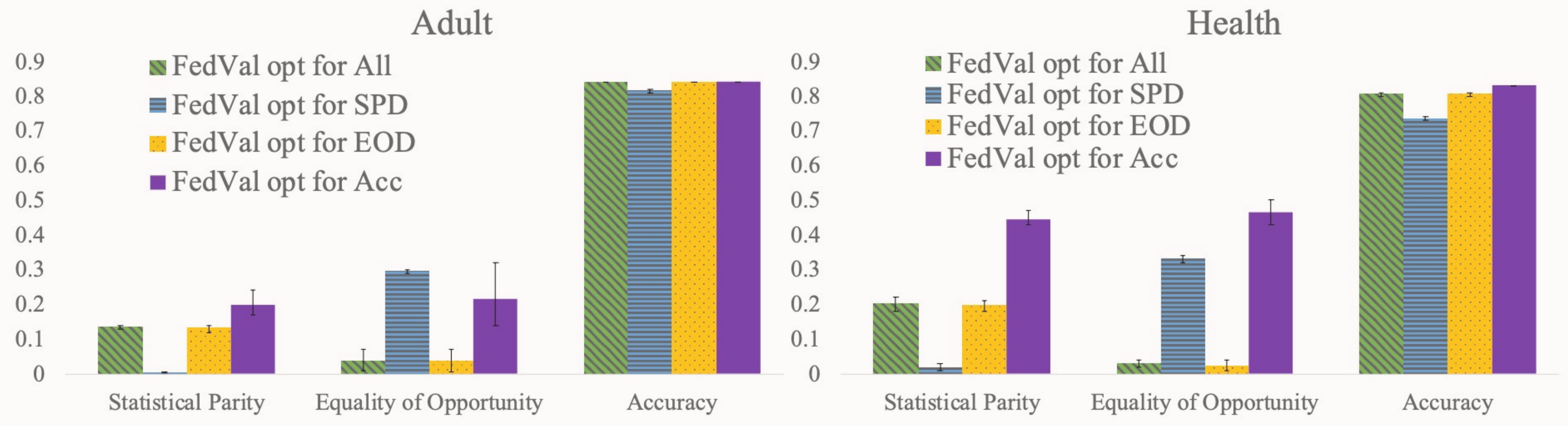
  return w to the server

Results

Verifying FedVal Against Baselines:



Verifying FedVal Against Different Objectives:



FedVal with Different Client Ratios: