FairFed: Enabling Group Fairness in Federated Learning (AAAI 2023)

SeHyun Park

March 21, 2024

Seoul National University

Outline

Introduction

Introduction

- With the increasing interest in Deep Learning and AI, the importance of fairness has come to the forefront. However, most fairness research assumes centralized learning where data is concentrated on main server.
- Apart from fairness, there's also a growing interest in data privacy, with Federated Learning(FL) being one method to address it.
- This paper proposes one approach to enhancing group fairness in the decentralized data setting of Federated Learning, named FairFed.

Related Work

- Federated Learning is a decentralized machine learning approach where models are trained collaboratively across multiple clients while keeping data localized.

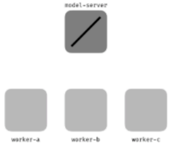
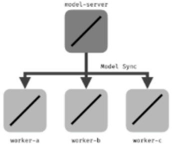
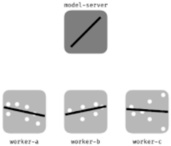
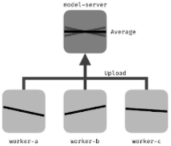
Step 1	Step 2	Step 3	Step 4
			
Central server chooses a statistical model to be trained	Central server transmits the initial model to several nodes	Nodes train the model locally with their own data	Central server pools model results and generate one global mode without accessing any data

Figure 1: Federated learning general process, https://en.wikipedia.org/wiki/Federated_learning

Related Work

- Fairness in Federated Learning
  1. *client-based fairness* (Li et al. 2019; Mohri, Sivek, and Suresh 2019)
     - aim to equalize model performance across different clients.
  2. *collaborative fairness* (Lyu et al. 2020; Wang et al. 2021)
     - aim to reward a highly-contributing participant with a better performing local model than is given to a low-contributing participant.

Related Work

3. *group fairness*
   - 3-1) (Zhang, Kou, and Wang 2020; Du et al. 2021; Galvez et al. 2021) try to distributively solve an optimization objective with fairness constraints,
     which requires each client to share the statistics of the sensitive attributes of its local dataset to the server.
   - 3-2) In (Zeng, Chen, and Lee 2021), an adaptation of the FairBatch debiasing algorithm (Roh et al. 2021) is proposed for FL where clients use FairBatch locally and the weights are updated through the server in each round.
   - 3-3) In (Papadaki et al. 2021), an algorithm is proposed to achieve minimax fairness in federated learning.

- Above methods, the server require each client to explicitly share the performance of the model on each subgroup separately.
  - ex. males with +ve outcomes, females with +ve outcomes, etc

FairFed: Fairness-Aware Aggregation in FL

Notation

- Federated Learning Setup
  - clients : $C \in \{1, \ldots, K\}$
  - local dataset of client k : $\mathcal{D}_k$ of size n_k
  - feature, label, and sensitive attribute : $(\mathbf{X}, Y, A) \in \mathcal{D}$
  - local objective of client k : $\mathcal{L}_k(\theta) = \frac{1}{n_k} \sum_{(\mathbf{X},Y,A) \in D_k} \ell(\theta, \mathbf{X}, Y)$

$\Rightarrow$ Find $\theta^* = \underset{\theta}{argmin} \sum_{k=1}^{K} w_k \mathcal{L}_k(\theta)$, where $w_k \geq 0, \sum w_k = 1$

$*$ In the federated averaging algorithm FedAvg, $w_k = \frac{n_k}{n}$, where $n = \sum_{k=1}^{K} n_k$

Notation

- **Group Fairness**

In this paper, they primarily consider *Equal Opportunity*.

- Global group fairness

  $F_{global} = Pr(\widehat{Y} = 1|A = 0, Y = 1) - Pr(\widehat{Y} = 1|A = 1, Y = 1)$

- Local group fairness

  $F_k = Pr(\widehat{Y} = 1|A = 0, Y = 1, C = k) - Pr(\widehat{Y} = 1|A = 1, Y = 1, C = k)$

FairFed: Fairness-Aware Aggregation in FL

- Clients are assumed to use their own in-processing methods to improve local group fairness.
- Recall that in the t-th round in FedAvg, local model updates $\{\theta_k^t\}_{k=1}^K$ are weight-averaged to get the new global model parameter θ^t as : $\theta^t = \sum_{k=1}^K w_k^t \theta_k^t$, where the weights $w_k^t = n_k / \sum_k n_k$

$\Rightarrow$ A fairness-oblivious aggregation would favor clients with more datapoints.

- In this paper, they propose a method to optimize F_{global} via adaptively adjusting the aggregation weights of different clients based on their F_k

FairFed: Fairness-Aware Aggregation in FL

- Computing Aggregation Weights for FairFed
  - At the beginning of training, $w_k^0 = n_k / \sum_{k=1}^{K} n_k$
  - The weight update follows this formular $\forall k \in \{1, \ldots, K\}$:

$$\Delta_k^t = \begin{cases} |Acc_k^t - \overline{Acc^t}| & \text{if } F_k^t \text{ is undefined} \\ |F_{global}^t - F_k^t| & \text{otherwise} \end{cases},$$

$$\bar{\omega}_k^t = \bar{\omega}_k^{t-1} - \beta \left(\Delta_k - \frac{1}{K} \sum_{i=1}^{K} \Delta_i \right), \quad \omega_k^t = \frac{\bar{\omega}_k^t}{\sum_{i=1}^{K} \bar{\omega}_i^t}.$$

where Acc_k^t and $\overline{Acc^t}$ represent the local accuracy at client k and global accuracy.

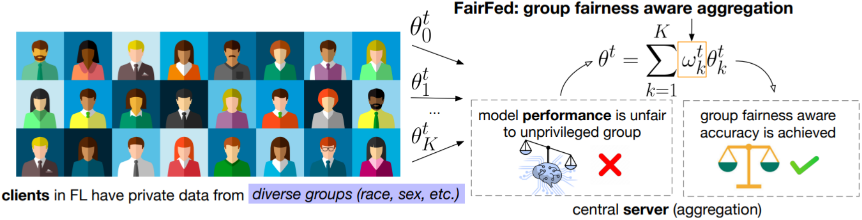
  - We need to compute F_{global}^t, while ensuring that raw local data is not directly transferred to server.

- Decompose F_{global} for the FL scenario.

$$
\begin{aligned}
F_{global} &= \Pr(\hat{Y}=1 \mid A=0, Y=1) - \Pr(\hat{Y}=1 \mid A=1, Y=1) \\
&= \sum_{k=1}^{K} \frac{n_k}{n} \underbrace{\left[\frac{\Pr(\hat{Y}=1 \mid A=0, Y=1, C=k) \Pr(A=0, Y=1 \mid C=k)}{\Pr(Y=1, A=0)} - \frac{\Pr(\hat{Y}=1 \mid A=1, Y=1, C=k) \Pr(A=1, Y=1 \mid C=k)}{\Pr(Y=1, A=1)} \right]}_{m_{global,k}}
\end{aligned}
$$

$\Rightarrow$ If each node knows the value of Since $m_{global,k}$ can be computed locally in each node, aggregate it using the **Secure Aggregation (SecAgg)** algorithm.

- Framework and Algorithm



Algorithm 1: FairFed Algorithm (tracking EOD)

Initialize: global model parameter θ_0 and weights $\{\omega_k^0\}$ as $\omega_k^0 = n_k / \sum_{i=1}^{K} n_i, \forall k \in [K]$;
Dataset statistics: Aggregate statistics $\mathcal{S} = \{\Pr(A{=}1, Y{=}1), \Pr(A{=}0, Y{=}1)\}$ from clients using Secure Aggregation (SecAgg) and send it to clients;
for *each round* $t = 1, 2, \cdots$ **do**
$F_{global}^t, \overline{Acc}^t \leftarrow \text{SecAgg}\left(\{\textbf{ClientLocalMetrics}\,(k, \theta^{t-1})\}_{k=1}^{K}\right)$; *// SecAgg to get* $\overline{Acc}^t$ *and global fairness* F_{global}^t *as in* (7);
$\frac{1}{K}\sum_i \Delta_i \leftarrow \text{SecAgg}\left(\{\textbf{ClientMetricGap}\,(k, \theta^{t-1}, F_{global}^t, \overline{Acc}^t)\}_{k=1}^{K}\right)$; *// SecAgg to compute mean of metric gaps*;
// Compute aggregation weights locally at clients based on (6) *then use SecAgg to aggregate weighted local model updates*;
$\left(\sum_{k=1}^{K} \bar{\omega}_k^t \theta_k^t\right), \left(\sum_{k=1}^{K} \bar{\omega}_k^t\right) \leftarrow \text{SecAgg}\left(\{\textbf{ClientWeightedModelUpdate}\,(k, \theta^{t-1}, \omega_k^t, \frac{1}{K}\sum_i \Delta_i)\}_{k=1}^{K}\right)$;
$\theta^{t+1} \leftarrow \left(\sum_{k=1}^{K} \bar{\omega}_k^t \theta_k^t\right) / \left(\sum_{k=1}^{K} \bar{\omega}_k^t\right)$;

Figure 2: FairFed framework and algorithm

Experiments

Setup

▶ **Dataset and sensitive attribute**

(1) Adult dataset : sex
(2) COMPAS dataset : race (white or non-white)

▶ **Configurable data heterogeneity for diverse sensitive attribute distributions**

- They draw $\mathbf{q}_k \sim \mathrm{Dir}(\alpha \mathbf{p})$ for each client k, where $\mathbf{p}$ represents the proportions of sensitive attributes and α be a hyperparameter.
- Then, the number of samples corresponding to $\mathbf{q}_k$ of the sensitive attribute is assigned to the client.

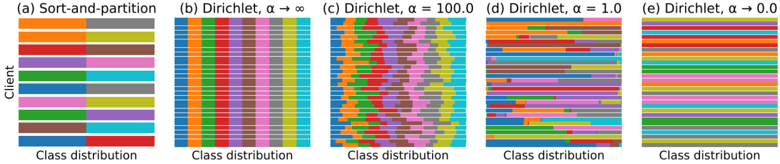


Figure 3: Synthetic populations with non-identical clients on CIFAR-10.

Setup

- Baseline
  - **FedAvg**: the original FL algorithm for distributed training of private data.
  - **FedAvg + Global reweighting [Global RW]** (Abay et al. 2020): A differential-privacy approach to collect noisy statistics such as the number of samples with privileged attribute values (A=1) and favorable labels (Y =1) from clients.
  - **FedFB** (Zeng, Chen, and Lee 2021): An in-processing debiasing approach in FL based on FairBatch.

▶ Result

	Method	Adult ($\beta = 1$)					COMPAS ($\beta = 1$)				
		Heterogeneity Level α					Heterogeneity Level α				
		0.1	0.2	0.5	10	5000	0.1	0.2	0.5	10	5000
Acc.	FedAvg	**0.835**	**0.836**	**0.835**	**0.836**	**0.837**	**0.674**	**0.673**	**0.675**	**0.674**	**0.675**
	Local / [Best]	0.831	0.833	0.834	0.831	0.829	0.666	0.659	0.665	0.663	0.664
	Global RW	0.834	0.833	0.831	0.829	0.829	0.673	0.671	0.672	0.676	0.675
	FedFB	0.825	0.825	0.829	0.832	0.832	0.674	0.673	0.675	0.677	0.677
	FairFed / RW	0.830	0.834	0.832	0.829	0.829	0.672	0.670	0.669	0.669	0.673
	FairFed / FairRep	0.824	0.833	0.834	0.834	0.834	0.661	0.655	0.663	0.663	0.660
	FairFed / FairBatch	0.829	0.833	0.830	0.830	0.831	0.659	0.664	0.665	0.661	0.661
EOD	FedAvg	-0.174	-0.173	-0.176	-0.179	-0.180	-0.065	-0.071	-0.067	-0.076	-0.078
	Local / [Best]	0.052	-0.009	-0.006	-0.013	0.014	-0.055	-0.051	-0.054	-0.038	-0.035
	Global RW	-0.030	0.019	0.022	0.017	0.010	-0.060	-0.065	-0.066	-0.076	-0.077
	FedFB	-0.019	0.015	0.015	-0.012	-0.012	-0.062	-0.061	-0.063	-0.077	-0.072
	FairFed / RW	**-0.017**	**0.001**	0.018	0.016	0.013	-0.057	-0.065	-0.053	-0.067	-0.061
	FairFed / FairRep	0.023	-0.009	-0.071	-0.174	-0.187	**0.037**	**0.023**	**0.043**	0.046	0.039
	FairFed / FairBatch	-0.020	**0.001**	**0.000**	**-0.005**	**-0.004**	-0.048	-0.048	-0.049	**-0.035**	**-0.031**

End