

Pursuing Overall Welfare in Federated Learning through Sequential Decision Making





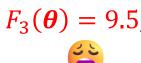
Seok-Ju Hahn, Gi-Soo Kim, Junghye Lee

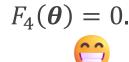
Client-Level Fairness in Federated Learning

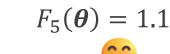
• Problem) Non-uniform performance distribution of a global model, θ

$$F_1(\boldsymbol{\theta}) = 0.1$$











• Simple Solution) Use of an adaptive mixing coefficient, p

$$\min_{\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^d} F(\boldsymbol{\theta}) \coloneqq \sum_{i=1}^K w_i F_i(\boldsymbol{\theta}), w_i = \frac{n_i}{\sum_{j=1}^K n_j} \rightarrow \sum_{i=1}^K p_i F_i(\boldsymbol{\theta}), \boldsymbol{p} \in \Delta_{K-1}$$

Imposing **LARGER** coefficients to local updates with **LARGER** losses

• Sample Deficiency) Only a single single response vector (e.g., local losses $[F_1(\theta), ..., F_K(\theta)]^{\mathsf{T}}$) is given for deciding a single mixing coefficient vector, $\mathbf{p} = [p_1, ..., p_K]^{\mathsf{T}}$...

Fair FL Algorithms are an Online Convex Optimization Algorithms

- Exponentiated Gradient
 - For all t=1,...,T, suppose we want to minimize a decision loss $\ell^{(t)}(\boldsymbol{p})=-\langle \boldsymbol{p},\boldsymbol{r}^{(t)}\rangle$ sequentially, for a response vector $m{r}^{(t)} \in \mathbb{R}^K$ and a decision variable $m{p} \in \Delta_{K-1}$:

$$p^{(t+1)} = \underset{\boldsymbol{p} \in \Delta_{K-1}}{\operatorname{argmin}} \ell^{(t)}(\boldsymbol{p}) + \eta R(\boldsymbol{p})$$

• When the regularizer $R(\boldsymbol{p}) = \sum_{i=1}^K p_i \log p_i$, a negative entropy, with a step size $\eta \in \mathbb{R}_{\geq 0}$, the closed form update is given as follows.

$$p_i^{(t+1)} = \frac{p_i^{(t)} \exp(r_i^{(t)}/\eta)}{\sum_{j=1}^{K} p_j^{(t)} \exp(r_j^{(t)}/\eta)}$$

| Method | Response, $r_i^{(t)}$ | Last Decision, $p_i^{(t)}$ | Step Size, η | New Decision, $p_i^{(t+1)}$ | | |
|-----------------------------------|--|----------------------------|-------------------|---|--|--|
| FedAvg | 0 | w_i | 1 | $\propto w_i$ | | |
| q-FedAvg (AFL if $q \to \infty$) | $q \log F_i(\boldsymbol{\theta}^{(t)})$ | w_i | 1 | $\propto w_i F_i^q(\boldsymbol{\theta}^{(t)})$ | | |
| TERM | $F_i(oldsymbol{	heta}^{(t)})$ | w_i | $1/\lambda$ | $\propto w_i \exp\left(\lambda F_i(\boldsymbol{\theta}^{(t)})\right)$ | | |
| PropFair | $-\log\left(M-F_i(\boldsymbol{\theta}^{(t)})\right)$ | w_i | 1 | $\propto \frac{w_i}{M - F_i(\boldsymbol{\theta}^{(t)})}$ | | |

Follow-The-Regularized-Leader (FTRL)

- OCO is certainly a suitable choice for making decision under the sample deficiency.
- Existing fair FL algorithms were NOT devised for online learning.
- Better decision-making is possible with the following modified objective.

$$p^{(t+1)} = \underset{p \in \Delta_{K-1}}{\operatorname{argmin}} \sum_{\tau=1}^{t} \ell^{(\tau)}(p) + \eta^{(t+1)} R(p) \left(\operatorname{or} R^{(t+1)}(p) \right)$$

AAggFF: Adaptive Aggregation for Fair Federated Learning

- AAggFF-S: Cross-Silo FL Setting
 - Number of clients (K) < Total communication rounds (T)
 - Full synchronization of clients

$$p^{(t+1)} = \underset{p \in \Delta_{K-1}}{\operatorname{argmin}} \sum_{\tau=1}^{t} \tilde{\ell}^{(\tau)}(p) + \frac{\alpha}{2} ||p||_{2}^{2} + \frac{\beta}{2} \sum_{\tau=1}^{t} (\langle g^{(\tau)}, p - p^{(\tau)} \rangle)^{2}$$

- Runtime: $O(K^2 + K^3)$ (cubic complexity due to weighted simplex projection)
- AAggFF-D: Cross-Device FL Setting
 - Number of clients $(K) \gg$ Total communication rounds (T)
 - Partial synchronization of clients (i.e., client sampling is required)

$$\boldsymbol{p}^{(t+1)} = \underset{\boldsymbol{p} \in \Delta_{K-1}}{\operatorname{argmin}} \sum_{\tau=1}^{t} \tilde{\ell}^{(\tau)}(\boldsymbol{p}) + \frac{L_{\infty}\sqrt{t+1}}{\sqrt{\log K}} \sum_{i=1}^{K} p_{i} \log p_{i}$$

- Runtime: $\mathcal{O}(K)$ (closed-form update exists)
- Doubly-Robust Estimator for Partially Observed Response (in the Cross-Device Setting)
 - Denote $C = P(i \in S^{(t)})$ as a client sampling probability,

 $S^{(t)}$ is an index set of selected clients, and $\bar{\mathbf{r}}^{(t)} = \frac{1}{|S^{(t)}|} \sum_{i \in S^{(t)}} r_i^{(t)}$. Then,

$$\check{r}_i^{(t)} = \left(1 - \frac{\mathbb{I}(i \in S^{(t)})}{C}\right)\bar{\mathbf{r}}^{(t)} + \frac{\mathbb{I}(i \in S^{(t)})}{C}r_i^{(t)},$$

which satisfies $\mathbb{E}[\breve{r}^{(t)}] = r^{(t)}$.

Vanishing Regret Guarantees

- Decision Objective (Regret Minimization): Regret $(T)(p^*) \triangleq \sum_{t=1}^T \ell^{(t)}(p^{(t)}) \sum_{t=1}^T \ell^{(t)}(p^*)$
- Decision Loss (Negative Logarithmic Growth): $\ell^{(t)}(p) \triangleq -\log(1+\langle p, r^{(t)}\rangle)$
- Theorem 1 (Regret of AAggFF-S): Regret $^{(T)}(p^*) \le 2L_{\infty}K\left(1 + \log\left(1 + \frac{T}{16K}\right)\right)$
- Theorem 2 (Regret of AAggFF-D w/o sampling): Regret $^{(T)}(p^*) \leq 2L_{\infty}\sqrt{T\log K}$
- \bigcirc Corollary 3 (Regret of AAggFF-D w/ sampling): $\mathbb{E}\left[\operatorname{Regret}^{(T)}(\boldsymbol{p}^{\star})\right] \leq 2L_{\infty}\sqrt{T\log K}$

Improved Uniformity in Performance Distribution

(Number of Clients: K / Number of Rounds: T)

| Dataset | | Berka (AUROC) | | | \mathbf{MQP} | | | ISIC (Acc. 5) | | | | |
|---|--------------|------------------|--------------|----------------|----------------|--------------|--------------|-----------------------|--------------|--------------|-----------------------|-----------------------|
| | | | | | (AUROC) | | | | | | | |
| | Avg. | Worst | Best | Gini | Avg. | Worst | Best | Gini | Avg. | Worst | Best | Gini |
| | (\uparrow) | (\uparrow) | (\uparrow) | (\downarrow) | (\uparrow) | (\uparrow) | (\uparrow) | (\downarrow) | (\uparrow) | (\uparrow) | (\uparrow) | (\downarrow) |
| -1 A | 80.09 | 48.06 | 99.03 | 10.87 | 56.06 | 41.03 | 76.31 | 8.63 | 87.42 | 69.92 | 92.57 | 4.84 |
| | (2.45) | (25.15) | (1.37) | (4.11) | (0.06) | (4.33) | (8.42) | (0.91) | (2.11) | (6.78) | (2.56) | (1.17) |
| V ET | 79.70 | 49.02 | 98.55 | 10.58 | 56.01 | 41.28 | 75.54 | 8.56 | 87.39 | 68.17 | 93.33 | 4.80 |
| $\begin{array}{c} {\tt AFL} \\ (4.14) \end{array}$ | (4.14) | (25.89) | (2.05) | (5.03) | (0.30) | (3.92) | (6.77) | (1.24) | (2.31) | (10.09) | (2.18) | (1.74) |
| - Fod Arra | 79.98 | 49.44 | 98.07 | 10.62 | 56.89 | 40.22 | 79.38 | 8.68 | 41.59 | 20.38 | 58.08 | 22.25 |
| q-FedAvg | (3.89) | (26.15) | (2.73) | (5.22) | (0.42) | (3.06) | (9.09) | (0.57) | (16.22) | (23.24) | (28.52) | (10.02) |
| TERM 80.11 | 80.11 | 48.96 | 99.03 | 10.86 | 56.47 | 40.73 | 76.80 | 8.67 | 87.89 | 77.32 | <u>96.00</u> | 3.77 |
| LNY | (3.08) | (25.79) | (1.37) | (4.73) | (0.19) | (4.36) | (8.30) | (1.43) | (1.69) | (5.84) | (3.27) | (0.94) |
| 79.24 FedMGDA | 79.24 | 46.38 | 99.03 | 11.64 | 53.02 | 34.91 | 69.65 | 10.33 | 42.36 | 21.44 | 59.21 | 22.25 |
| edrigda | (2.96) | (24.11) | (1.37) | (4.84) | (1.67) | (2.22) | (3.89) | (0.44) | (14.94) | (21.30) | (28.52) | (10.02) |
| DwanEair | 79.61 | 49.44 | 98.07 | 10.47 | 56.60 | 41.71 | <u>79.09</u> | 8.74 | 83.88 | 58.36 | 91.35 | 7.91 |
| PropFair (4 | (4.49) | (26.15) | (2.73) | (5.04) | (0.39) | (3.80) | (7.40) | (0.87) | (2.50) | (11.63) | (2.48) | (2.10) |
| AggFF-S | 80.93 | 52.08 | 99.03 | 10.16 | <u>56.63</u> | 41.79 | 75.56 | 8.38 | 89.76 | 85.17 | $\boldsymbol{98.22}$ | 2.52 |
| (Proposed) | (2.96) | (23.59) | (1.37) | (3.80) | (0.54) | (4.43) | (6.53) | (0.77) | (1.03) | (3.87) | (1.66) | (0.38) |

| Dataset | CelebA (Acc. 1) | | | | Reddit (Acc. 1) | | | | | | | | |
|--------------|------------------------|---------------------|------------------|----------------|-----------------|------------------|------------------|----------------|--------------|------------------|------------------|----------------|--|
| Method | | | | | | | | | | | | | |
| | Avg. | Worst | Best | Gini | Avg. | Worst | Best | Gini | Avg. | Worst | Best | Gini | |
| | (\uparrow) | $10\% \ (\uparrow)$ | $10\%(\uparrow)$ | (\downarrow) | (\uparrow) | $10\%(\uparrow)$ | $10\%(\uparrow)$ | (\downarrow) | (\uparrow) | $10\%(\uparrow)$ | $10\%(\uparrow)$ | (\downarrow) | |
| FedAvg | 90.79 | 55.76 | 100.00 | 7.86 | 10.76 | 2.50 | 20.86 | 25.66 | <u>75.51</u> | 7.93 | 100.00 | 24.58 | |
| | (0.53) | (0.84) | (0.00) | (0.30) | (1.45) | (0.21) | (3.64) | (0.49) | (1.08) | (2.87) | (0.00) | (1.34) | |
| q-FedAvg | 90.88 | 55.73 | <u>100.00</u> | 7.82 | 12.76 | 3.38 | 21.81 | 23.34 | 73.34 | 11.19 | <u>100.00</u> | 23.16 | |
| | (0.19) | (0.85) | (0.00) | (0.21) | (0.32) | (0.20) | (0.19) | (0.34) | (0.47) | (0.47) | (0.00) | (0.13) | |
| TERM | 90.71 | 55.66 | 100.00 | 7.90 | 12.02 | 2.85 | 20.74 | 24.15 | 70.90 | 5.98 | 100.00 | 26.37 | |
| | (0.65) | (0.93) | (0.00) | (0.38) | (0.16) | (0.41) | (0.65) | (1.05) | (2.96) | (1.10) | (0.00) | (1.32) | |
| FedMGDA | 88.33 | 48.60 | 100.00 | 9.75 | 10.58 | 2.35 | 19.09 | 25.20 | 72.45 | 9.65 | 100.00 | 23.68 | |
| | (0.63) | (25.85) | (0.00) | (0.59) | (0.18) | (0.20) | (0.62) | (0.22) | (1.88) | (2.90) | (0.00) | (1.27) | |
| PropFair | 87.25 | 48.11 | 100.00 | 10.39 | 11.26 | 1.95 | 21.33 | 25.97 | 73.64 | 7.30 | 100.00 | 24.97 | |
| | (5.01) | (10.03) | (0.00) | (3.43) | (0.71) | (0.32) | (0.92) | (1.02) | (3.31) | (1.02) | (0.00) | (1.09) | |
| AAggFF-D | 91.27 | 56.71 | 100.00 | 7.54 | 12.95 | 4.75 | 22.81 | 22.59 | 76.68 | 14.54 | 100.00 | 21.42 | |
| (Duan aga 1) | (0.07) | (0.00) | (0.00) | (0.04) | (0.00) | (0.70) | (1.00) | (0.00) | (0.00) | (0 =0) | (0.00) | (0.01) | |

■ Cross-Device Setting $(K \gg T)$

Setting (K < T)

