Pursuing Overall Welfare in Federated Learning through Sequential Decision Making

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

ICML International Conference On Machine Learning

UNIST ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY 2007

Client-Level Fairness in Federated Learning

- Problem) Non-uniform performance distribution of a global model, $\boldsymbol{\theta}$

$F_1(\boldsymbol{\theta}) = 0.1,\quad F_2(\boldsymbol{\theta}) = 2.3,\quad F_3(\boldsymbol{\theta}) = 9.5,\quad F_4(\boldsymbol{\theta}) = 0.6,\quad F_5(\boldsymbol{\theta}) = 1.1$

!! VIOLATION of the CLIENT-LEVEL FAIRNESS !!

- Simple Solution) Use of an adaptive mixing coefficient, $\boldsymbol{p}$

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

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

- Sample Deficiency) Only a single single response vector (e.g., local losses $[F_1(\boldsymbol{\theta}), \dots, F_K(\boldsymbol{\theta})]^\top$) is given for deciding a single mixing coefficient vector, $\boldsymbol{p} = [p_1, \dots, p_K]^\top$…

Fair FL Algorithms are an Online Convex Optimization Algorithms

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

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

  - 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.

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

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	$F_i(\boldsymbol{\theta}^{(t)})$	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.

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

AAggFF: **A**daptive **Agg**regation for **F**air **F**ederated Learning

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

$$\boldsymbol{p}^{(t+1)} = \operatorname*{argmin}_{\boldsymbol{p}\in\Delta_{K-1}} \sum_{\tau=1}^{t} \tilde{\ell}^{(\tau)}(\boldsymbol{p}) + \frac{\alpha}{2}\|\boldsymbol{p}\|_2^2 + \frac{\beta}{2}\sum_{\tau=1}^{t}\left(\langle \boldsymbol{g}^{(\tau)}, \boldsymbol{p} - \boldsymbol{p}^{(\tau)}\rangle\right)^2$$

  - Runtime: $\mathcal{O}(K^2 + K^3)$ (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)

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

  - 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 $\bar{\mathrm{r}}^{(t)} = \frac{1}{|S^{(t)}|}\sum_{i\in S^{(t)}} r_i^{(t)}$. Then,

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

which satisfies $\mathbb{E}\left[\check{\boldsymbol{r}}^{(t)}\right] = \boldsymbol{r}^{(t)}$.

Vanishing Regret Guarantees

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

Improved Uniformity in Performance Distribution

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

◀ Cross-Silo Setting ($K < T$)

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

◀ Cross-Device Setting ($K \gg T$)

Dataset	CelebA (Acc. 1)				Reddit (Acc. 1)				SpeechCommands (Acc. 5)			
Method	Avg. (↑)	Worst 10% (↑)	Best 10%(↑)	Gini (↓)	Avg. (↑)	Worst 10%(↑)	Best 10%(↑)	Gini (↓)	Avg. (↑)	Worst 10%(↑)	Best 10%(↑)	Gini (↓)
FedAvg	90.79 (0.53)	55.76 (0.84)	100.00 (0.00)	7.86 (0.30)	10.76 (1.45)	2.50 (0.21)	20.86 (3.64)	25.66 (0.49)	75.51 (1.08)	7.93 (2.87)	100.00 (0.00)	24.58 (1.34)
q-FedAvg	90.88 (0.19)	55.73 (0.85)	100.00 (0.00)	7.82 (0.21)	12.76 (0.32)	3.38 (0.20)	21.81 (0.19)	23.34 (0.34)	73.34 (0.47)	11.19 (0.47)	100.00 (0.00)	23.16 (0.13)
TERM	90.71 (0.65)	55.66 (0.93)	100.00 (0.00)	7.90 (0.38)	12.02 (0.16)	2.85 (0.41)	20.74 (0.65)	24.15 (1.05)	70.90 (2.96)	5.98 (1.10)	100.00 (0.00)	26.37 (1.32)
FedMGDA	88.33 (0.63)	48.60 (25.85)	100.00 (0.00)	9.75 (0.59)	10.58 (0.18)	2.35 (0.20)	19.09 (0.62)	25.20 (0.22)	72.45 (1.88)	9.65 (2.90)	100.00 (0.00)	23.68 (1.27)
PropFair	87.25 (5.01)	48.11 (10.03)	100.00 (0.00)	10.39 (3.43)	11.26 (0.71)	1.95 (0.32)	21.33 (0.92)	25.97 (1.02)	73.64 (3.31)	7.30 (1.02)	100.00 (0.00)	24.97 (1.09)
AAggFF-D (Proposed)	**91.27** (0.07)	**56.71** (0.08)	100.00 (0.00)	**7.54** (0.04)	**12.95** (0.39)	**4.75** (0.76)	**22.81** (1.36)	**22.59** (0.28)	**76.68** (0.80)	**14.54** (2.58)	100.00 (0.00)	**21.42** (0.81)