Ditto: Fair and Robust Federated Learning Through Personalization

Tian Li (CMU)  Shengyuan Hu (CMU)  Ahmad Beirami (Facebook AI)  Virginia Smith (CMU)
Motivation

constraints in federated learning

- **fairness**
- **robustness**
- privacy
- security
- communication

... ...

representation disparity

against data and model poisoning attacks

competing with each other

\[ w^* \in \arg \min_w G \left( F_1(w), \ldots, F_K(w) \right) \]
Insights

**personalization** to achieve robustness and fairness simultaneously

for each device $k \in [K]$,

Ditto:

$$
\min_{v_k} h_k(v_k; w^*) := F_k(v_k) + \frac{\lambda}{2} \|v_k - w^*\|^2
$$

s.t. $w^* \in \arg\min_w G(F_1(w), \ldots, F_K(w))$

* simple form of MTL: ensure personalized models are close to global model
* easy to implement in federated settings
* accurate, robust, and fair
Setup

Robustness: Byzantine robustness
- (A1) label poisoning: flipped, or random noisy labels
- (A2) random Gaussian updates
- (A3) model replacement

measurement: mean test performance across benign devices

Fairness: representation disparity*

measurement: test performance deviation across benign devices

*Fairness without Demographics in Repeated Loss Minimization, Hashimoto et al, ICML 2018
Ditto: analyze robustness/fairness

We first look at a simplified federated point estimation problem:

local objective function: \( \min_{v_k} F_k(v_k) = \frac{1}{2} \left( v_k - \frac{1}{n} \sum_{i=1}^{n} x_{k,i} \right)^2 \)

- \( X_1 = \{x_{1,1}, \ldots, x_{1,n}\} \) number of local samples
- \( \tau \) task unrelatedness
- \( K_a \) number of malicious devices
- \( \tau_a \) strength of the attack
- \( v_1 \sim \mathcal{N}(0, \sigma^2) \)
- \( v_k \sim \mathcal{N}(0, \sigma^2) \)
- \( v_j \sim \mathcal{N}(0, \tau^2) \)

\( \theta \) malicious devices
Ditto: analyze robustness/fairness

explicitly characterize the form of $\lambda^*$:

$$\lambda^* = \frac{\sigma^2}{n} + \frac{K}{K^2 + \frac{K_a}{K-1} (\tau_a^2 - \tau^2)}$$

- test accuracy and variance are jointly minimized with $\lambda^*$
- $n \to \infty \implies \lambda^* \to 0$
- $K_a \to \infty$ or $\tau_a \to \infty \implies \lambda^* \to 0$
- $K_a = 0$, $\tau$ increases $\implies \lambda^*$ decreases
- $\tau = 0$, $\tau_a > \tau \implies \lambda^* < \infty$

$n$: number of local samples
$\tau$: task unrelatedness
$K_a$: number of malicious devices
$\tau_a$: strength of the attack
Ditto: analyze robustness/fairness

All these results can be generalized to a class of linear problems.
Ditto Solver

+ personalization add-on

**Algorithm 1: Ditto for Personalized FL**

1. **Input**: $K$, $T$, $s$, $\lambda$, $\eta$, $w^0$, $\{v_k^0\}_{k \in [K]}$
2. for $t = 0, \cdots, T - 1$ do
3. Server randomly selects a subset of devices $S_t$, and sends the current global model $w^t$ to them
4. for device $k \in S_t$ in parallel do
5. Solve the local sub-problem of $G(\cdot)$ inexactely starting from $w^t$ to obtain $w_k^t$:
   \[ w_k^t \leftarrow \text{UPDATE\_GLOBAL}(w^t, \nabla F_k(w^t)) \]
   /* Solve $h_k(v_k; w^t)$ */
6. Update $v_k$ for $s$ local iterations:
   \[ v_k = v_k - \eta (\nabla F_k(v_k) + \lambda(v_k - w^t)) \]
   Send $\Delta_k^t := w_k^t - w^t$ back
7. Server aggregates $\{\Delta_k^t\}$:
   \[ w^{t+1} \leftarrow \text{AGGREGATE}(w^t, \{\Delta_k^t\}_{k \in [S_t]}) \]
8. **return** $\{v_k\}_{k \in [K]}$ (personalized models), $w^T$ (global model)

- a scalable, simple personalization add-on for any federated global solver
- preserves the practical properties of the global FL solver (e.g., communication, privacy)
- with convergence guarantees
Modularity of Ditto

* **Optimization:** can plug in any global model solver, and inherit the convergence benefits
  
  [Theorem] If $w^*$ converges with rate $g(t)$, then there exists $c < \infty$ such that Ditto converges with rate $cg(t)$

* **Privacy:** Ditto preserves privacy/communication benefits of the global objective and its solver

* **Robustness:** can plug in existing robust aggregators to robustify $w^*$
Experiments

fair methods are not robust

robust methods are not fair (with high variance)
Experiments

Ditto is more robust than strong baselines under various attacks.

Ditto is also more fair on average, improve absolute accuracy by $\sim 6\%$ over the strongest robust baseline and reduce variance by $\sim 10\%$ over SOTA fair methods.
Future Work

- Do other personalization formulations offer similar benefits?
- What is the optimal personalization formulation for FL?
- Can we further characterize the effect of personalization in terms of fairness, robustness, privacy, etc?
Ditto: Fair and Robust Federated Learning Through Personalization

Full Paper: https://arxiv.org/abs/2012.04221

Thanks!