Ditto: Fair and Robust Federated Learning Through Personalization

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

ICML | 2021

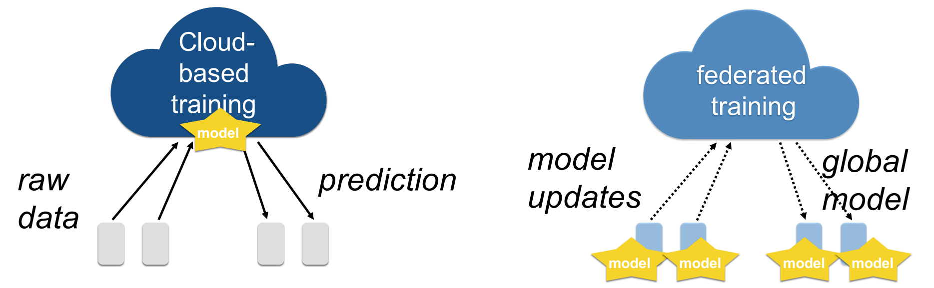
Motivation

Pragmatic constraints in federated learning: fairness, robustness, privacy, security, etc.

Simultaneously satisfying these (competing) constraints can be exceptionally difficult

This work: constraints between **accuracy**, **fairness** (performance uniformity), and **robustness** (against data and model poisoning attacks)*

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



* Fairness: the uniformity of performance distribution
Robustness: the average test accuracy, across benign devices.

Ideas

Properly modeling *statistical heterogeneity*

Method: federated multi-task learning

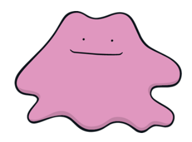
- A simple and effective multi-task learning objective for personalization federated learning (and a scalable solver with convergence guarantees)

Theoretically and empirically, *we show that*

- Personalization (Ditto) can offer inherent robustness and fairness
- Personalization (Ditto) is particularly useful to handle multiple constraints simultaneously

Global-Regularized Federated Multi-Task Learning

Objective



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

$$\min_{v_k} \quad h_k(v_k; w^*) := F_k(v_k) + \frac{\lambda}{2}\|v_k - w^*\|^2$$
$$\text{s.t.} \quad w^* \in \arg\min_w G\left(F_1(w), \ldots, F_K(w)\right)$$

Enforce personalized models to be close to w^*

w^* is the optimal global model

Solver

At each round, first randomly sample a subset of devices S_t. For each device in S_t:

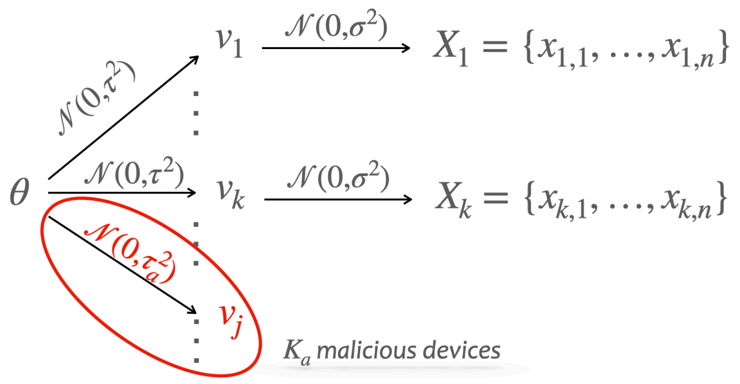
$w_k^t \leftarrow$ UPDATE_GLOBAL $\left(w^t, \nabla F_k(w^t)\right)$, $\Delta_k^t := w_k^t - w^t$

$v_k = v_k - \eta\left(\nabla F_k(v_k) + \lambda(v_k - w^t)\right)$ ← Ditto add-on

Server: $w^{t+1} \leftarrow$ AGGREGATE $\left(w^t, \{\Delta_k^t\}_{k \in \{S_t\}}\right)$

- A scalable, simple personalization add-on for any federated global solver
- Preserves the practical properties of the global solver (communication, privacy)
- With convergence guarantees

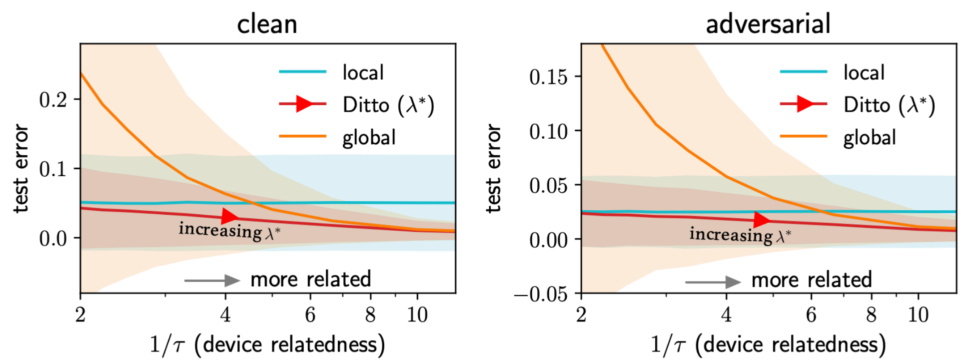
Analysis of Ditto in Simplified Settings



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

τ: task unrelatedness; τ_a: strength of the attack

- Test accuracy and variance are jointly minimized with λ^*
- $n \to \infty \implies \lambda^* \to 0$
- $K_a \to \infty$ or $\tau_a \to \infty \implies \lambda^* \to 0$
- $K_a = 0$, τ increases $\implies \lambda^*$ decreases

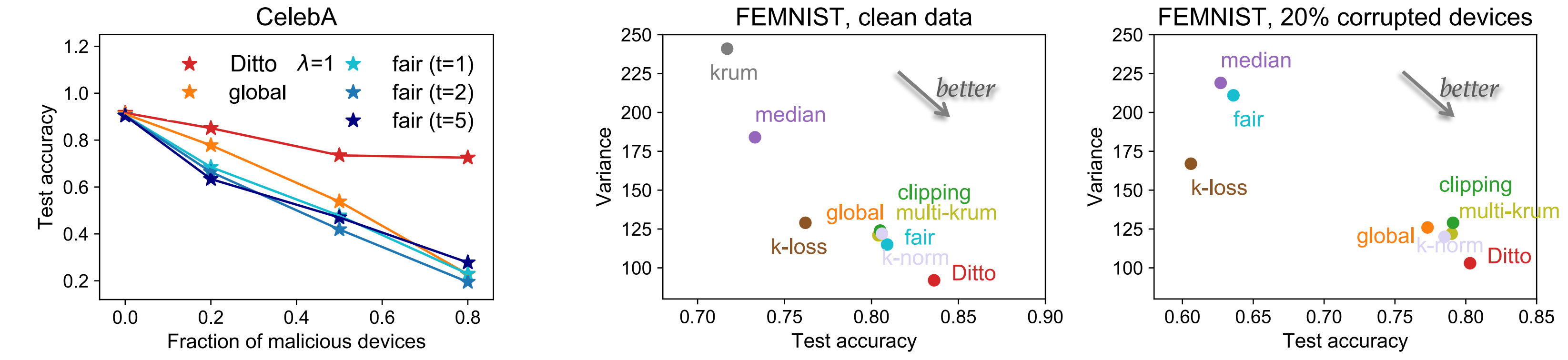


Results hold for linear problems

(i) Ditto is superior than learning global or local models
(ii) λ^* should increase as the increase of device relatedness $(1/\tau)$

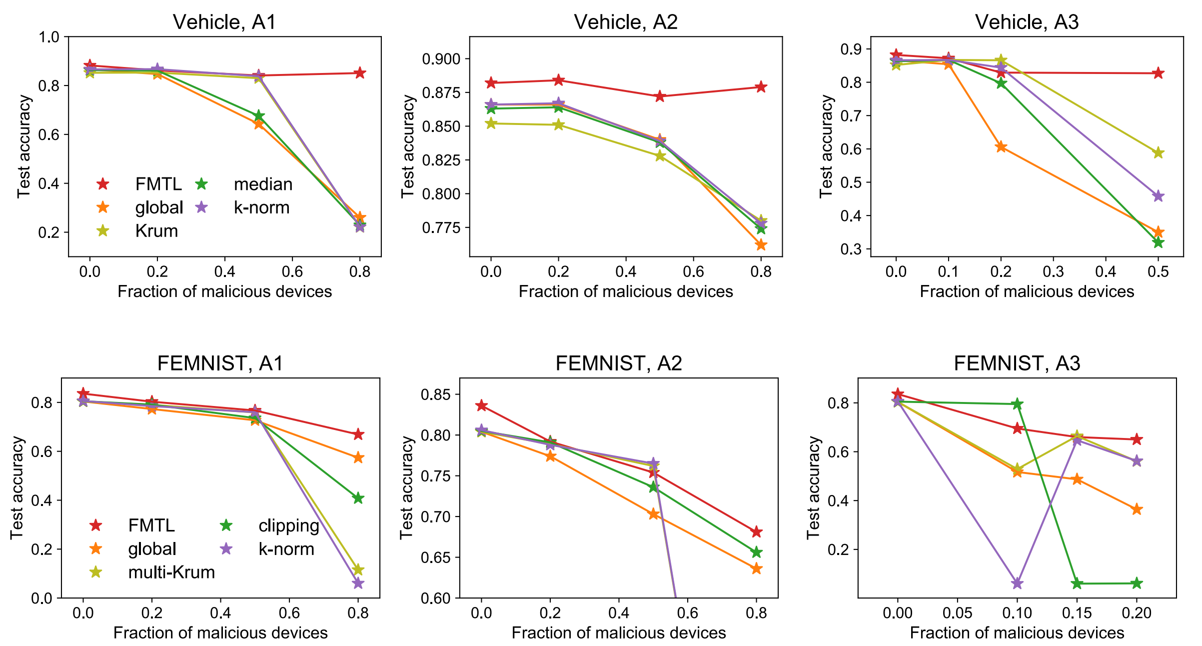
Evaluation

LEAF: A Benchmark for Learning in Federated Settings



Fair methods are not robust

Robust methods are not fair (with high variance)
Ditto is both robust and fair



Ditto is more robust than strong baselines under various attacks

A1: data corruption
A2: sending random Gaussian updates
A3: data corruption + model replacement

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?

Code: https://github.com/litian96/ditto **ArXiv:** https://arxiv.org/pdf/2012.04221.pdf