Federated Multi-Task Learning for Competing Constraints

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

NeurIPS-SpicyFL 2020

Motivation

pragmatic constraints in federated learning: fairness, robustness, privacy, security, …

Simultaneously satisfying these constraints can be exceptionally difficult

This work: constraints between **accuracy**, **fairness** (performance uniformity), and **robustness** (to training-time attacks)*

They are competing constraints in statistically heterogeneous networks.

- Fairness method: susceptible to attacks from malicious devices
- Robust method: can be unfair to different devices

** We define fairness as the uniformity of performance distribution, and robustness as the average test accuracy, across benign devices.*

Insights

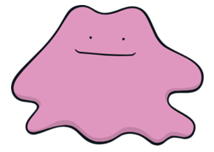
Key idea: properly modeling *statistical heterogeneity*

Method: federated multi-task learning

Contributions:

- a simple yet effective multi-task learning objective to achieve robustness and fairness in a unified framework (with a lightweight solver)
- first theoretical analysis on the benefits of MTL to fairness and robustness (on a toy problem)
- experiments on diverse attacks across a set of federated datasets —> the proposed objective is both fair and robust

Ditto: Federated Multi-Task Learning Objective for Competing Fairness and Robustness Constraints



$$\min_{\{w_k\},1\le k\le N} \sum_{k=1}^{N} p_k \left(F_k(w_k) + \frac{\lambda}{2}\|w_k - w^*\|^2\right),$$

Enforce personalized models to be close to w^*

$$\text{s.t. } w^* = \arg\min_{w} \sum_{k=1}^{N} p_k F_k(w)$$

w^* is the optimal global model

At each round t:

$$w_k^t = w_k^t - \eta \nabla F_k(w_k^t),$$
$$v_k = v_k - \eta\left(\nabla F_k(v_k) + \lambda(v_k - w_k^t)\right)$$

each selected device $k \in S_t$ run local updates

$$w^{t+1} = w^t + \frac{1}{|S_t|}\sum_{k\in S_t} \Delta_k^t$$

server aggregates global model updates

- lightweight, scalable, easy to optimize in federated settings
- theoretically motivated (see example below)
- outperforms fairness and robustness baselines
- also achieves state-of-the-art (or higher) accuracy in terms of personalization

Analysis of Fairness-Robustness Tradeoffs

Setup:

federated point estimation: $f_k(w) = \frac{1}{2}\left\| w - \frac{1}{n}\sum_{i\in[n]} x_{k,i} \right\|_2^2$, and

$\theta \rightarrow \xrightarrow{\mathcal{N}(0,\tau^2)}$ model on device k : $w_k \xrightarrow{\mathcal{N}(0,\sigma^2)}$ data point $x_{k,i}$

✓ <u>No adversaries: Ditto for accuracy and fairness</u>

$$\textbf{Optimal } \lambda = \frac{\sigma^2}{\boxed{n}\boxed{\tau^2}}$$

number of local samples

device relatedness

✓ <u>With adversaries: Ditto for accuracy, fairness, and robustness</u>

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

number of malicious devices

capability of malicious devices

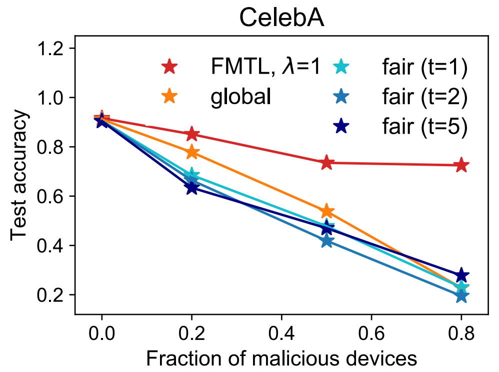
(τ_a^2: variance of malicious w_k)

explicitly characterize the benefits of Ditto with respect to various constraints:

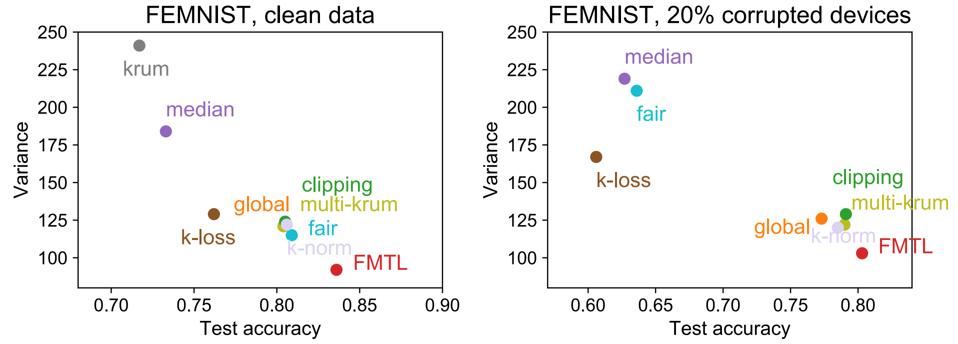
optimal λ has a closed-form, which is a function of number of local data points, device relatedness, number of malicious devices, capability of malicious devices, etc.

Evaluation

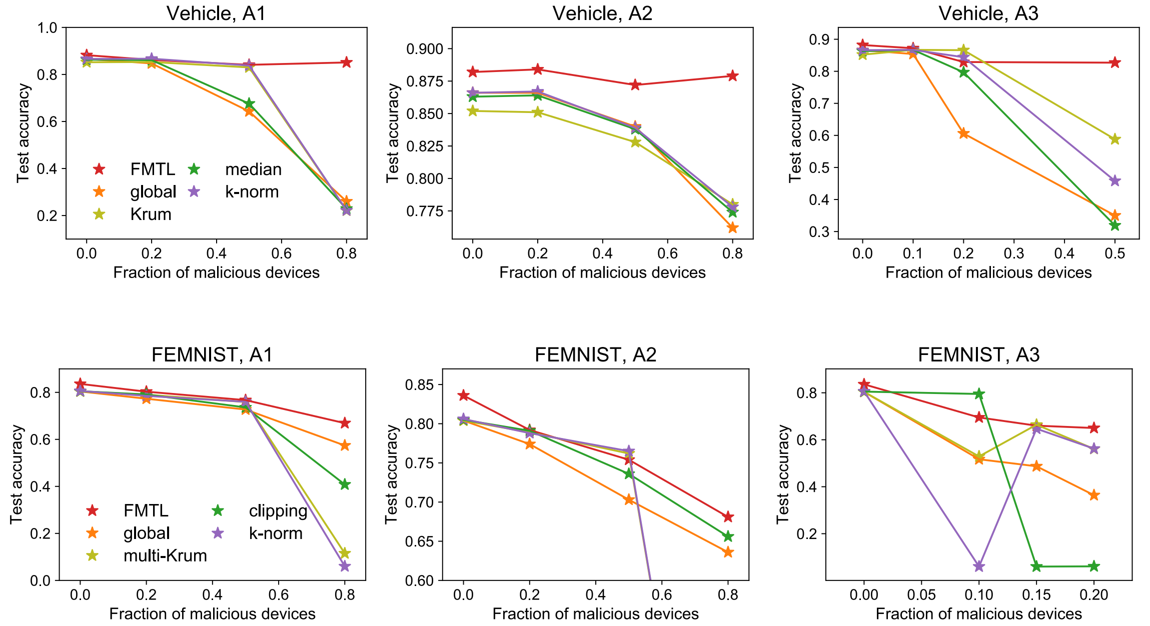
LEAF: A Benchmark for Learning in Federated Settings (leaf.cmu.edu)



fair methods are not robust



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



Ditto (FMTL) is more robust than strong baselines under various attacks

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

Future Work

- Understand the interplay between other constrains (i.e., privacy and fairness, privacy and robustness), for other notions of fairness and robustness?
- Further theoretical understanding on the benefits of MTL for fairness, robustness, and privacy?
- Optimal MTL framework to handle these constraints? Other schemes (i.e., clustering)?