Robust Decentralized Federated Learning Against both Distributional Shifts and Byzantine Attacks

Kaichuang Zhang[1], Ping Xu[1], Zhi Tian[2]
1 University of Texas Rio Grande Valley, 2 George Mason

Abstract

Decentralized federated learning (DFL) offers enhanced resilience to client failures and potential attacks compared to its centralized counterpart.

Nonetheless, the adoption of DFL in practical applications faces several challenges that can threaten the **robustness** of local models. On one hand, **distributional shifts**, the distribution of data might change over time or space due to the decentralized infrastructure, which impacts the aggregated model's performance on both local agents and test data. On the other hand, **Byzantine attacks**, which involve certain users sending malicious updates to their neighbors to propagate erroneous knowledge, could compromise the convergence and accuracy of the global model.

We observe that there has not yet been any work that claims to resolve both spatial and temporal distributional shifts, let alone when combined with Byzantine attacks. **In this work, we aim to fill this gap by proposing a robust decentralized learning algorithm that is resilient to both distributional shifts and Byzantine attacks.**

Problem Statement

Distributionally Robust Optimization under DFL with Byzantine Attacks:

$$w^* = \arg\min_{w} \sup_{Q\in\Omega} \frac{1}{|\mathcal{B}|}\sum_{k\in\mathcal{B}}\frac{1}{|Q_k|}\sum_{\xi\in Q_k} f_k(w,\xi) \quad (1)$$

where
$\mathcal{W}$ is the learning model;
Q is worst-case in Ω;
Ω is ambiguity set which constructed based on the information of empirical distribution of training dataset $\mathcal{D}$;
$\mathcal{B}$ is the benign clients in DFL system;
Q_k is the dataset of client k and a part of Q;
f_k is the loss function of client k;
ξ is the data sample in Q_k;

Proposed Algorithm

Robust Aggregation Rule

NBS: filter out model updates with large norms. In this way, the influence of a malicious update is properly bounded: it either is filtered out for having a large norm or can only finitely impact the output with a norm comparable to some benign updates. Detailed in **Algorithm 1.**

Algorithm 1 Norm-based Screening at client n

Input: $\boldsymbol{X}_n = \{\boldsymbol{x}_n^{t+\frac{1}{2}}, \boldsymbol{x}_m^{t+\frac{1}{2}}(\forall m \in \mathcal{N}_n)\}$, that consists of client n's local model $\boldsymbol{x}_n^{t+\frac{1}{2}}$, updates from neighbors $\boldsymbol{x}_m^{t+\frac{1}{2}}, \forall m \in \mathcal{N}_n$, screening percentage β
Output: x_n^{t+1}

1: generate a new set of indices $(1),\ldots,(m)$, such that $\|x_{(1)}^{\frac{1}{2}}\| \le \cdots \le \|x_{(m)}^{\frac{1}{2}}\|$
2: define an index set $\mathcal{U} = \{(1),\ldots,((1-\beta)m)\}$, which specifies the unscreened inputs
3: calculate the output by averaging the unscreened inputs $x_n^{t+1} = \frac{1}{|\mathcal{U}|}\sum_{i\in\mathcal{U}} \boldsymbol{x}_i^{t+\frac{1}{2}}$

Distributional & Byzantine Robust Distributed Gradient

NBS against Byzantine attacks;
Wasserstein DRO against distributional shifts.
Ω in Eq.(1) will use Wasserstein distance.

$$W_c(Q,\mathcal{D}) \le \rho$$

[1] proves Eq.(1) can be soloved use

$$\arg\min_{w}\sup_{z}\{f(\theta;z) - \lambda c(z,x)\}$$

Algorithm 2 Distributional & Byzantine Robust Distributed Gradient Descent at client n

Input: screening percentage $\beta(\ge\alpha)$, learning rate η, model initialization $\boldsymbol{x}_n^0$, total iteration T
Output: completed model x_n^T

1: **for** $t = 0, 1, \ldots, T-1$ **do**
2: **for** $i = 1, 2, \ldots, m$ **do**
3: obtain $z_j^{\epsilon}(\theta_t)$ for each local sample x_j by solving $\sup_z\{f(\theta_t;z) - \lambda c(z,x_j)\}$ to ϵ-precision
4: local training

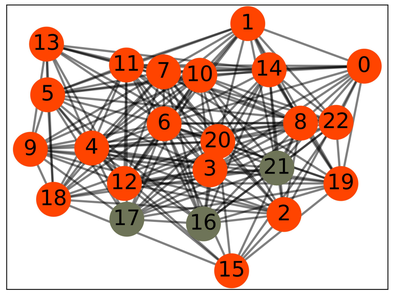
$$\boldsymbol{x}_n^{t+\frac{1}{2}} = \begin{cases} \star & n \in \mathcal{M} \\ \boldsymbol{x}_n^{t+\frac{1}{2}} = \boldsymbol{x}_n^t - \eta^t \nabla f_n(\boldsymbol{x}_n^t;\xi_n^t), & n \in \mathcal{B} \end{cases} \quad (12)$$

5: send $\boldsymbol{x}_n^{t+\frac{1}{2}}$ to client n' neighbors
6: compute the aggregated model $x_n^{t+1} = \text{Norm\_Screen}_\beta(x_1(\theta_t),\ldots,x_m(\theta_t))$
7: **end for**
8: **end for**

Simulation Results

Network

20 benign nodes 3 Byzantine nodes



Byzantine Attacks : Sign-Flipping

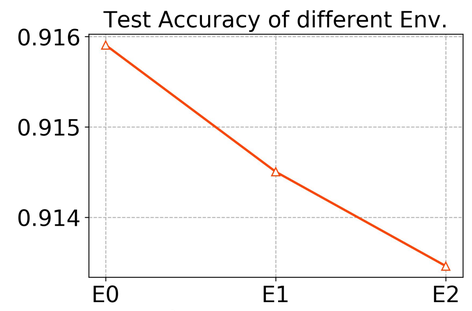
$$x_n^{t+\frac{1}{2}} = * = -10 \times \frac{1}{|\mathcal{B}_n|}\sum_{x_i^{t+\frac{1}{2}}\in\mathcal{B}_n} x_i^{t+\frac{1}{2}}$$

Distributional Shifts

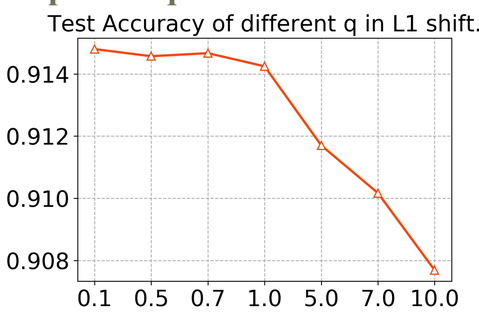
L1 shift: $|z-x| \le q$ L2 shift: $|z-x|_2 \le q$

Env	DS	BA
E0	0	No-Attack
E1	L1 with $q=0.3$	Sign-Flipping
E2	L2 with $q=0.3$	Sign-Flipping

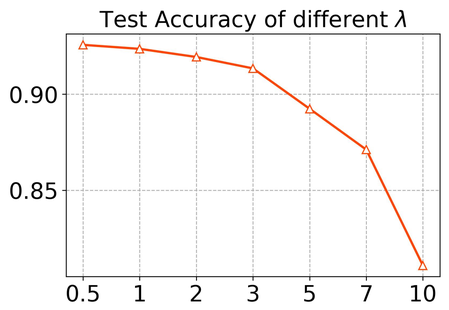
Different environments

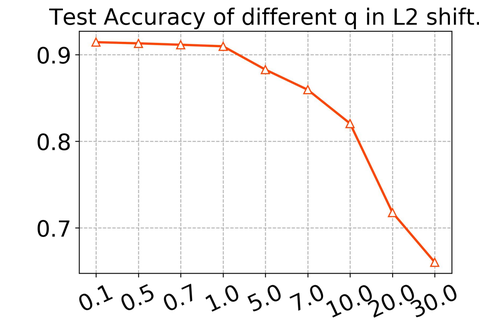


Impact of q:



Impact of λ:





Conclusions

In this work, we address the uncharted problem of robust decentralized federated learning at the presence of both distributional shifts and byzantine attacks. We propose a new algorithm that incorporates effective robust features to defend against both safety threats. We also empirically demonstrate that our algorithm enjoys satisfying performance.

Bibliography

[1]. Sinha A, Namkoong H, Volpi R, et al. Certifying some distributional robustness with principled adversarial training[J]. arXiv preprint arXiv:1710.10571, 2017.
[2]. Zhou G, Xu P, Wang Y, et al. Robust distributed learning against both distributi-onal shifts and byzantine attacks[J]. arXiv preprint arXiv:2210.16682, 2022.