VANDERBILT UNIVERSITY

Byzantine Resilient Distributed Multi-Task Learning

Jiani Li, Waseem Abbas and Xenofon Koutsoukos

Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN

VANDERBILT UNIVERSITY INSTITUTE FOR SOFTWARE INTEGRATED SYSTEMS

Abstract

The work studies **resilient distributed multi-task learning (MTL)** problem in a network of agents aiming to learn distinct but correlated models simultaneously. Some of the agents in the network are possibly **Byzantine** and their objective is to disrupt and prevent normal (non-adversarial) agents from achieving the MTL objective. We present and analyze a Byzantine-resilient distributed MTL approach whereby a normal agent **adaptively assigns weights** to its neighbors based on **similarities** with them, which are measured based on the accumulated loss of the **agent's data and its neighbors' models**. A small accumulated loss indicates a large similarity and vice-versa. We show that using the proposed weight assignment, normal agents with convex models **converge resiliently** towards their true target.

Highlights

- The proposed approach is resilient to an arbitrary number of Byzantine agents and require no knowledge of the number of Byzantine agents in the network.
- The learning performance is guaranteed to be at least as good as the non-cooperative case.
- A normal agent computes weights in time that is linear in the size of its neighborhood and the dimension of the data.

MTL

- A network of m agents $G = (V, E)$.
- Each agent k has data $\{(x_k^i, y_k^i)\}$ (sampled randomly from the distribution of ξ_k.)
- Prediction function: $f_k(x_k^i) = \theta_k^T x_k^i$
- Model parameter: θ_k
- Loss function: $\ell_k(.)$
- Expected risk function: $r_k(\theta_k) = \mathbb{E}[\ell_k(\theta_k; \xi_k)]$

Objective:

$$\min_{\Theta} \left\{ \sum_{k=1}^{m} r_k(\theta_k) + \eta \mathcal{R}(\Theta, \Omega) \right\},$$

where, $\Theta = [\theta_1, \theta_2, \ldots, \theta_m]$, $\mathcal{R}(.)$ is a regularization function, and Ω is a relationship matrix among agents.

Distributed MTL

Adapt-then-Combine Strategy: At each iteration i, agent k minimizes the individual risk using stochastic gradient descent (SGD) given local data followed by a combination step that aggregates neighboring models according to the weights assigned.

$$\hat{\theta}_{k,i} = \theta_{k,i-1} - \mu_k \nabla \ell_k(\theta_{k,i-1}; \xi_k^{i-1}) \qquad \text{(Adapt)}$$

$$\theta_{k,i} = \sum_{l \in \mathcal{N}_k} a_{lk} \hat{\theta}_{l,i} \quad \text{subject to} \quad \sum_{l \in \mathcal{N}_k} a_{lk} = 1, a_{lk} \geq 0, a_{lk} = 0 \text{ if } l \notin \mathcal{N} \qquad \text{(Combine)}$$

Main challenge is to design suitable **weights, a_{lk}**.

Resilient Distributed MTL

Goal: Design resilient online weight assignment rule for MTL in the presence of Byzantine agents.

Weight Assignment:

$$a_{lk}(i) = \begin{cases} \dfrac{r_k(\hat{\theta}_{l,i}^{(\text{coop})})^{-1}}{\sum_{p \in \mathcal{N}_k^{\leq}} r_k(\hat{\theta}_{p,i}^{(\text{coop})})^{-1}}, & \text{if } r_k(\hat{\theta}_{l,i}^{(\text{coop})}) \leq r_k(\hat{\theta}_{k,i}^{(\text{coop})}), \\ 0, & \text{otherwise,} \end{cases}$$

Main Results

Resilient convergence: Using the computed weights, each normal agent resiliently converges to the true target.

$$\lim_{i \to \infty} \theta_{k,i}^{(\text{coop})} = \theta_k^*, \quad \forall k,$$

Improved expected regret compared to non-cooperation: The computed weights result in an improved expected regret as compared to using the SGD without cooperation, even with Byzantine agents.

$$\lim_{i \to \infty} \sup R_k^{(\text{coop})}(i) \leq \lim_{i \to \infty} \sup R_k^{(\text{ncop})}(i), \forall k.$$

Computational efficiency: At each i, agent k computes weights in time linear in the size of neighborhood of k and the dimension of data.

Evaluation

We evaluated the resilience of the proposed online weight adjustment rule using three distributed MTL case studies, with and without Byzantine agents.

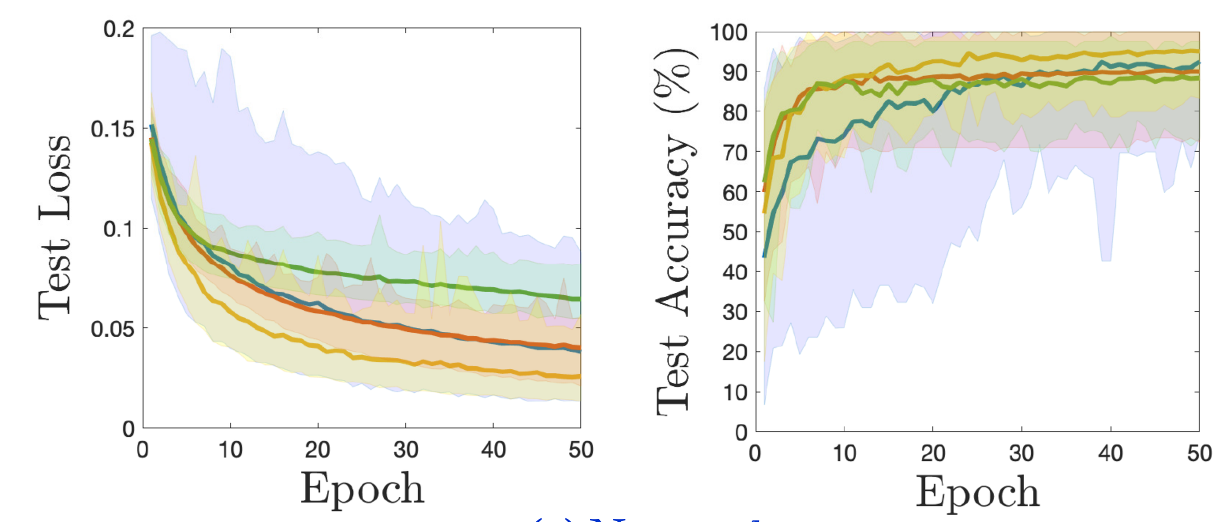
- Target localization (regression problem)
- Human activity recognition (classification)
- Digit classification (non-convex model)

Evaluation – Human Activity Recognition

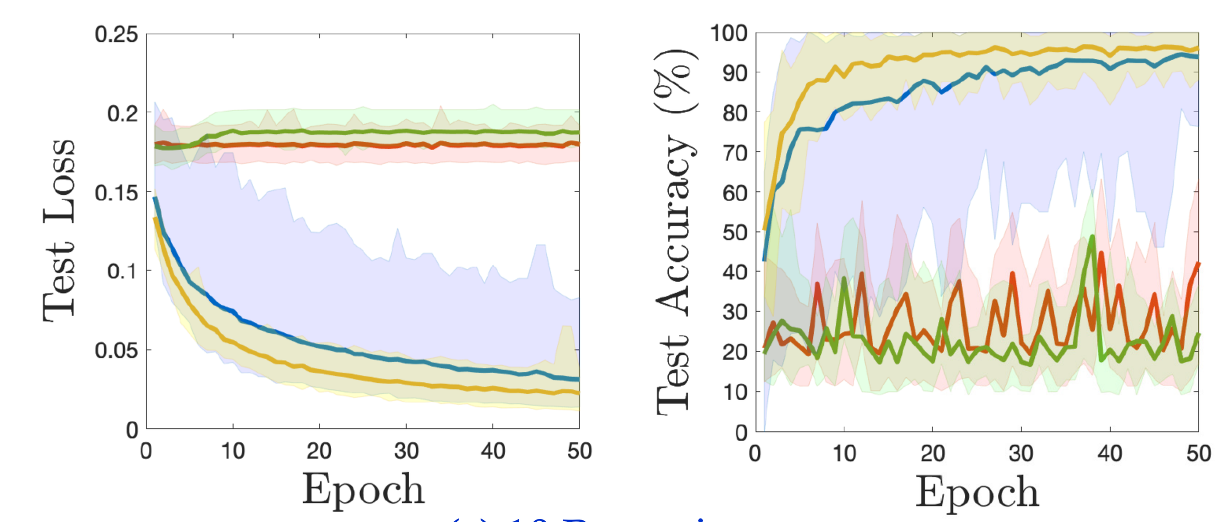
Goal: Predict activities (walking, walking upstairs, downstairs, sitting, standing, lying) using feature vectors generated by the sensor signals.

Set-up:

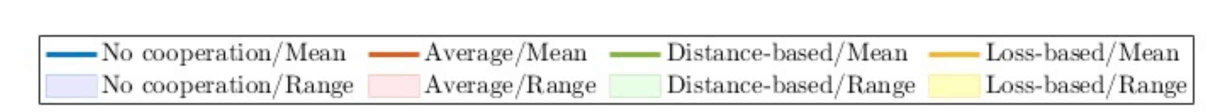
- Mobile phone sensor data of 30 individuals.
- Complete network topology.
- Byzantine agents continuously send random values.



(a) No attack



(c) 10 Byzantine agents



Concluding remark: The approach is easily extendable to general resilient distributed machine learning and federated learning systems.

34th Conference on Neural Information Processing Systems, 2020

NEURAL INFORMATION PROCESSING SYSTEMS