FedTop: A Constraint-Loosed Federated Learning Aggregation Method against Poisoning Attack

Che WANG, Zhenhao WU, Jianbo GAO,Jiashuo ZHANG,Junjie XIA,Feng GAO,Zhi GUAN,Zhong CHEN

Frontiers of Computer Science, DOI: 10.1007/s11704-024-3767-z

Problems & Ideas

- Problems of Federated Learning Aggregation Method:
  - Federated Learning is vulnerable to poisoning attacks that local clients could be manipulated by attacker.
  - Existing resistence methods tend to fail when facing severe poisoning attacks or have strong preconditions.

- Ideas: A novel performance-based method named FedTop, which is flexible to defend against poisoning attacks with less constraints.

Algorithm 1 FedTop

Input: Global round T; Initialized model parameters g^0; Number of total participants N; Number of submitted local models in each round m; Number of local models remained in each round $\hat{m}$

Output: Model parameters $g^{(T)}$

Initialization: $t = 0$
Server broadcasts $g^{(0)}$ to all participants
while $t < T$ **do**
  for $i \in [1, m]$ **do**
    Update local gradients $\nabla F_i\left(g^{(t)}; \mathcal{D}_i\right)$ on the data stored in client i
    Update local model: $l_i^{(t)} \leftarrow g^{(t)} - \Gamma \nabla F_i\left(g^{(t)}; \mathcal{D}_i\right)$
    Upload local model $l_i^{(t)}$ to the server;
  end for
  Transform all $l_i^{(t)}$ into normalized format $L_i^{(t)} = \frac{\|g^{(t)}\|}{\|l_i^{(t)}\|} l_i^{(t)}$
  Obtain the performance $p_i^{'}$ for every normalized local model $L_i^{(t)}$ on evaluation dataset $\mathcal{D}_g$
  Drop the last $m - \hat{m}$ normalized models with low performance scores $p_i^{'}$
  Update the weight of the remaining local model $p_i \leftarrow \frac{p_i^{'}}{\sum_{i=1}^{\hat{m}} p_i^{'}}$
  Aggregate the remaining local models into the new global model $g^{(t+1)} \leftarrow \sum_{i=1}^{\hat{m}} p_i \frac{\|g^{(t)}\|}{\|l_i^{(t)}\|} l_i^{(t)}$
  $t \leftarrow t + 1$
end while

The algorithm in the left describes the workflow of FedTop.

Main Contributions

- Contributions:
  - a novel poisoning attack defense method FedTop, which has better performance compared with existing methods;
  - A theoretical convergence analysis to prove the effectiveness and robustness of FedTop in IID, non-IID, and malicious environments;
  - Various experiments to verify the theoretical analysis that FedTop can defend against severe poisoning attacks effectively with high robustness;

Aggregation	MNIST		CIFAR-10		YELP	
methods	IID	Non-IID	IID	Non-IID	IID	Non-IID
FedAvg [6]	1.550	**1.560**	1.283	1.342	**1.081**	1.156
Median [3]	1.561	1.586	1.282	1.307	1.099	1.120
Multi-Krum [7]	1.584	1.659	1.304	1.386	1.120	1.775
Zeno [5]	1.552	1.567	1.324	1.394	1.102	1.182
FedTop	**1.548**	1.564	**0.943**	**1.069**	1.089	**1.118**

Aggregation	Scale Up			Mix			Scale Down		
methods	3	8	12	3	8	12	3	8	12
FedAvg [6]	2.021	2.056	2.045	1.968	2.027	2.033	1.779	2.078	2.235
Median [3]	1.695	2.046	2.089	1.657	1.998	2.085	1.656	2.245	2.250
Multi-Krum [7]	1.702	2.136	2.129	1.713	2.010	2.116	1.725	2.260	2.268
Zeno [5]	1.610	2.062	2.055	1.620	1.817	2.024	1.851	2.238	2.279
FedTop	**1.597**	**1.576**	**1.624**	**1.598**	**1.585**	**1.579**	**1.581**	**1.589**	**1.602**

Performance of FedTop compared with different methods under both normal and malicious environments. Left: the loss of different methods in normal environment; Right: the loss of different methods in malicious environment.