Federated Model Distillation with Noise-Free Differential Privacy 2021

IJCAI Lichao Sun , Lingjuan Lyu

Contributions

1. Centralized differential privacy(CDP) requires a central trusted party
2. Local differential privacy(LDP) only support shallow models such as logisitic regression and only focus on simple tasks and datasets
3. Conventional Fl system suffers from several intrinsic limitations:(1) it requires every party to share their local model weights in each round, thus limiting only to models with homogeneous architectures; (2) sharing model weight incurs a significant privacy issue of local model, as it opens all the internal state of the model to white-box inference attacks; (3) model weight is usually of much higher dimension than model predictions, resulting in huge communication overhead and higher privacy cost.

- Author propose FEDMD-NFDP, a novel federated model distillation framework with the new proposed noise-free differential privacy (NFDP) mechanism that guarantees each party's privacy without explicitly adding any noise.

- prove that NFDP with both replacement and without replacement sampling strategies can inherently ensure (ε, σ)-differential privacy, eliminating noise addition and privacy cost explosion issues explicitly in previous works.
- Extensive experiments on benchmark datasets, various settings (IID and Non-IID data distribution), and heterogeneous model architectures, demonstrate that F ED MDNFDP achieves comparable utility with only a few private samples that are randomly sampled from each party, validating the numerous benefits of our framework.

Method

- FEDMD-NFDP

  (1)during initialization phase, every party i updates its local model weights ω_i on a randomly sampled subset $(X_i, Y_i) \in D_i$ from local private training data D_i for T_i times without any collaboration

  (2)during collaboration phase, parties share the knowledge of their local model via their predictions on a subset of public data, X_p

Collaboration phase

$Y_p[0] = f_{\mathsf{Aggreg}}(\{Y_p^{i\in[N]}[0]\})$ ▷ Initial aggregation at the server
for $t \in [R]$ communication rounds **do**
  Server randomly samples a public subset $X_p[t+1] \in D_p$
  for $i \in [N]$ parties **do** ▷ Each party updates local weight w_i in parallel
    for $j \in [T_2]$ epochs **do**
      Digest: $w_i \leftarrow \textsc{Train}\ (w_i, X_p[t], Y_p[t])$
    end for
    for $j \in [T_3]$ epochs **do**
      Revisit: $w_i \leftarrow \textsc{Train}\ (w_i, X_i, Y_i)$
    end for
    Send $Y_p^i[t+1] = \textsc{Predict}(w_i; X_p[t+1])$ to the server
  end for
  $Y_p[t+1] = f_{\mathsf{Aggreg}}(\{Y_p^{i\in[N]}[t+1]\})$ ▷ Prediction aggregation at the server
end for

- NFDP mechanism

Theorem 1. *[NFDP mechanism: (ϵ, δ)-differential privacy of sampling without replacement] Given a training dataset of size n, sampling without replacement achieves $(\ln \frac{n+1}{n+1-k}, \frac{k}{n})$-differential privacy, where k is the subsample size.*

Theorem 2. *[NFDP mechanism: (ϵ, δ)-differential privacy of sampling with replacement] Given a training dataset of size n, sampling with replacement achieves $((k \ln \frac{n+1}{n}, 1 - \left(\frac{n-1}{n}\right)^k)$-differential privacy, where k is the subsample size.*

Lemma 1. *Algorithm 1 using sampling with replacement is consistently more private than using sampling without replacement for any $n > 0$ and $0 < k \leq n$.*

Experimental

- MNIST/FEDMNIST and CIFAR-10/CIFAR-100
  (1)IID - Non-IID
  (2)Local model - two or three-layer DNN
  (3)pytorch GPU NVIDIA Tesla V100 - N = 10

- subset of size 5000, R = 20,T1 = 20,T2 = 2,T3 = 1
- FEDMD-NFDP - FEDMD-NP/Centralized/FEDMD-LDP
- DP - (ε, σ)/model convergence/distillation approaches/IID Non-IID/number of party/

FEMNIST	k	ϵ	δ	Accuracy
FEDMD-NP	2880	$+\infty$	1	96.15%
Centralized	2880	$+\infty$	1	98.00%
FEDMD-NFDP	16	0.0027	0.0062	80.64%
FEDMD-NFDP	60	0.0090	0.0206	88.06%
FEDMD-NFDP	300	0.0452	0.0989	93.56%
FEDMD-NFDP	2880	0.4342	0.6321	96.63%

CIFAR-10	k	ϵ	δ	Accuracy
FedMD-NP	300	$+\infty$	1	86.88%
Centralized	300	$+\infty$	1	88.83%
FEDMD-NFDP	16	0.0260	0.0583	74.40%
FEDMD-NFDP	60	0.0867	0.1815	81.58%
FEDMD-NFDP	120	0.1734	0.3301	83.57%
FEDMD-NFDP	300	0.4336	0.6327	87.38%

Table 2: Comparisons with All Baselines