Locally Differentially Private Distributed Deep Learning via Knowledge Distillation（2022）

arxiv Di Zhuang, Mingchen Li and J. Morris Chang, Senior Member, IEEE

Contributions

1. A fully trusted aggregator barely exists
2. Less efficient and not scalable
3. At least two non-colluding honest data users might not be practical

- A novel, effective and efficient privacy-preserving distributed deep learning framework using LDP and Knowledge-Distillation.
- An active sampling approach to efficiently reduce the total number of queries from the data user to each data owners, so that to reduce the total cost of privacy budget.

Method

- Each data owner perturbs the query data's soft label (using LDP techniques)
- LDP -> piecewise Mechanism

Algorithm 2: Piecewise Mechanism for One-Dimensional Numerical Data (PM-ONE) [13]

Input: tuple $z_i \in [-1, 1]$; privacy budget ϵ.
Output: perturbed tuple $z_i' \in [-\Delta, \Delta]$.

$\Delta \longleftarrow \frac{e^{\epsilon/2}+1}{e^{\epsilon/2}-1}$;
$L(z_i) \longleftarrow \frac{\Delta+1}{2} \cdot z_i - \frac{\Delta-1}{2}$;
$R(z_i) \longleftarrow L(z_i) + \Delta - 1$;
Sample value v uniformly at random from $[0, 1]$;
if $v < \frac{e^{\epsilon/2}}{e^{\epsilon/2}+1}$ **then**
    Sample z_i' uniformly at random from $[L(z_i), R(z_i)]$;
else
    Sample z_i' uniformly at random from $[-\Delta, L(z_i)] \cup [R(z_i), \Delta]$;
return z_i'.

Algorithm 3: Piecewise Mechanism for Multidimensional Numerical Data (PM) [13]

Input: tuple $z \in [-1, 1]^k$; privacy budget ϵ.
Output: perturbed tuple $z' \in [-k \cdot \Delta, k \cdot \Delta]^k$.

$z' \longleftarrow < 0, 0, \ldots, 0 >$;
$m \longleftarrow max\{1, min\{k, \lfloor \frac{\epsilon}{2.5} \rfloor\}\}$;
Sample m values uniformly without replacement from $\{1, 2, \ldots, k\}$;
for *each sampled attribute* j **do**
  $z'_j = \frac{k}{m} \cdot PM\text{-}ONE(z_j, \frac{\epsilon}{m})$;
return z'.

- Knowledge Distillation -> Active query sampling

1) Select an initial subset of S unlabeled public data X^Q uniformly at random from X^P. Update $X^P \leftarrow X^P - X^Q$.
2) Use X^Q to query the teacher models, and use the distilled knowledge to train the student initial student model M_S.
3) For each available public data $x_i \in X_P$, evaluate it on the student model M_S. Let P_{ij} denote the probability of x_i belonging to class $j \in \{1, 2, \ldots, k\}$ predicted by M_S. Let $P_i = \{P_{i1}, P_{i2}, \ldots, P_{ik}\}$, and suppose $\sum_{l=1}^{k} P_{il} = 1$. Let P_i^* be the largest value (posterior probability) in P_i. Then, repeat the procedure below for S times to select S query samples:

$$\begin{aligned} x_i &\leftarrow X^P \\ P_i &\leftarrow M_S(x_i) \\ X^Q &\leftarrow X^Q \cup \underset{x_i}{argmin} \frac{1}{m-1} \sum_{l=1}^{k} (P_i^* - P_{il}) \\ X^P &\leftarrow X^P - X^Q \end{aligned} \tag{5}$$

Then, use X^Q to query the teacher models, and use the distilled knowledge to train the student initial student model M_S.
4) Repeat 3), until the student model meet the performance requirement or no more public data available (i.e., $X_P = \emptyset$).

Experimental

- CIFAR-10/MNIST/Fashion-MNIST with LDP -> Piecewise mechanism /Duchi's mechanism /Laplace mechanism
- LDP-DL -> (SOTA) DP-SGD/PATE/DP-FL

Datasets	CIFAR10 [16]		MNIST [17]		FashionMNIST [18]	
Approaches	Accuracy	Privacy Budget	Accuracy	Privacy Budget	Accuracy	Privacy Budget
LDP-DL	77.5%	5	98.1%	5	83.4%	5
	79.7%	8	98.8%	8	85.7%	8
DP-SGD [15]	73.0%	8	97.00%	8	-	-
PATE [10]	73.6%	5	97.7%	5	81.5%	5
	76.0%	8	98.2%	8	84.7%	8
DP-FL [12]	75.9%	5	96.4%	5	82.6%	5
	78.7%	8	97.2%	8	83.6%	8

TABLE 1: In Comparison with Existing Approaches.