Compressive Differentially-Private Federated Learning Through Universal Vector Quantization

Saba Amiri[1]; Adam Belloum[1]; Sander Klous[2]; Leon Gommans[3]

[1]MNS Group, University of Amsterdam

[2]CCI Group, University of Amsterdam

[3]Air France KLM

ABSTRACT

Federated machine learning is an essential vehicle for achieving privacy preserving machine learning. However, a federated learning mechanism is usually hindered by the overhead communication costs between the orchestrator and participants. Furthermore, refraining from sharing the data will not lead us to absolute privacy, e.g. in presence of privacy attacks. Differential Privacy has provided a set of rigorous privacy standards to protect individual records of a dataset being used by a randomized mechanism and could potentially mitigate some of the relevant privacy concerns. However, addition of differential privacy usually costs even more communication overhead, putting more pressure on uplink and downlink channels. In this work, we present a novel algorithm for achieving both differential privacy and reduced communication overhead through compression of client-server communication by means of quantization.

CONTACT

Saba Amiri
MNS Group, University of Amsterdam
1098 XH Amsterdam, The Netherlands
Email: s.amiri@uva.nl

Acknowledgement

This research has been performed as part of the **Enabling Personalized Intervention** (EPI) project. The EPI project is funded by the Dutch Science Foundation in the Commit2Data program (Grant Number: 628.011.028).

References

[1] Zamir, Ram, and Meir Feder. "On universal quantization by randomized uniform/lattice quantizers." IEEE Transactions on Information Theory 38.2 (1992): 428-436.
[2] Zamir, Ram, and Meir Feder. "On lattice quantization noise." IEEE Transactions on Information Theory 42.4 (1996): 1152-1159.
[3] Gariby, Tal, and Uri Erez. "On general lattice quantization noise." 2008 IEEE International Symposium on Information Theory. IEEE, 2008.
[4] Balle, Borja, and Yu-Xiang Wang. "Improving the Gaussian mechanism for differential privacy: Analytical calibration and optimal denoising." *International Conference on Machine Learning*. PMLR, 2018.

Problem Definition

- In many application domains such as healthcare, centralized machine learning is not feasible due to different problems, e.g.:
  - Inability to share data due to **data privacy**.
  - Large datasets containing "big data" are hard or impossible to **communicate**.

Current Solutions

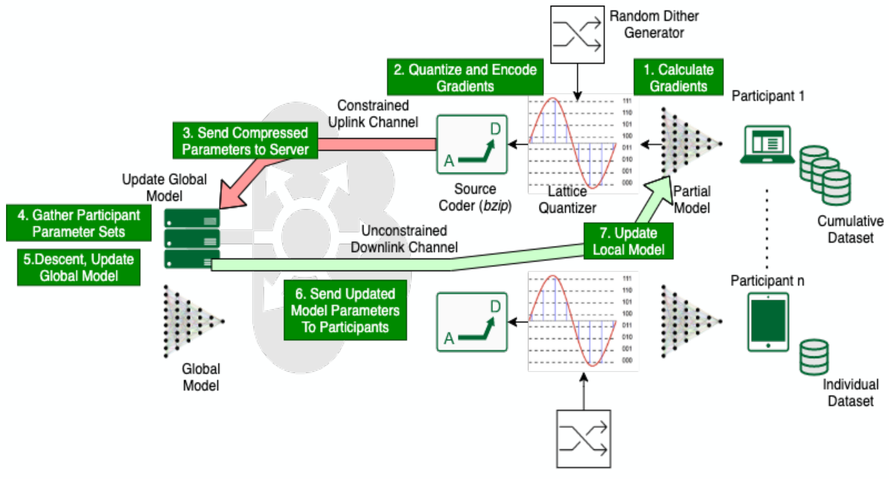
- **Federated Learning (FL)** for **communication/data sharing**
  - Federated learning solves the problem of data sharing, computation is brough to participants.
  - Causes **communication overhead**. This overhead is undesirable in the presence of constrained uplink/downlink channels.
- **Differential Privacy (DP)** for **privacy**
  - Adds a level of privacy to the machine learning pipeline.
  - DP is usually achieved by introducing perturbation to the system.

Research Question

- This work is motivated by
  - The need to introduce a certain level of privacy to our machine learning pipeline.
  - The necessity of compressing communications between participants and central authority, reduce the load on communication channels.
- With regards to the previously defined general problems of federated learning, namely communication overhead and privacy, we pose the research question "*Is it possible to use the perturbation resulting from compression of communication in a federated learning scenario to achieve Differential Privacy?*".

Our Solution

- **Quantization:** We use dithered lattice quantization[1] followed by universal source coding as the compression method. Problem with Lattice Quantization: Noise is correlated with source distribution.
  - **Dither** is a vector of i.i.d random variable. Addition of dither to the quantization mechanism ensures independence of quantization noise from source distribution
  - **Universal Source Coding**: after performing quantization, we use universal source coding mechanism of *bzip*, creating a codebook and performing the compression step.
- **Differential Privacy**
  - We aim to achieve local differential privacy on sample level
  - We use quantization noise as the DP perturbation mechanism
  - We shape the quantizer noise[2,3] to adhere to Gaussian PDF $N(0, \sigma I)$, adhering to Analytical Gaussian Mechanism[4] constraints for $\varepsilon > 1$. The parameter σ is determined based on our target privacy budget and quantization levels.



- System diagram for our proposed compressive differentially private federated learning method. The uplink channel is constrained. Seven steps of each learning round are marked on the chart.

Algorithms

Algorithm 1: Compressive (ϵ, δ)-DP Federated Learning - Participant i, Round $t > 1$

Input: Training samples $D_i = \{x_{1i}, ..., x_{Ni}\}$, loss function $\mathcal{L}(\theta) = \frac{1}{N_i} \sum_j \mathcal{L}(\theta, x_{ij})$
Parameters: gradient norm bound S_q, group size G, lattice vector size L, quantization step size Δ, noise scale N

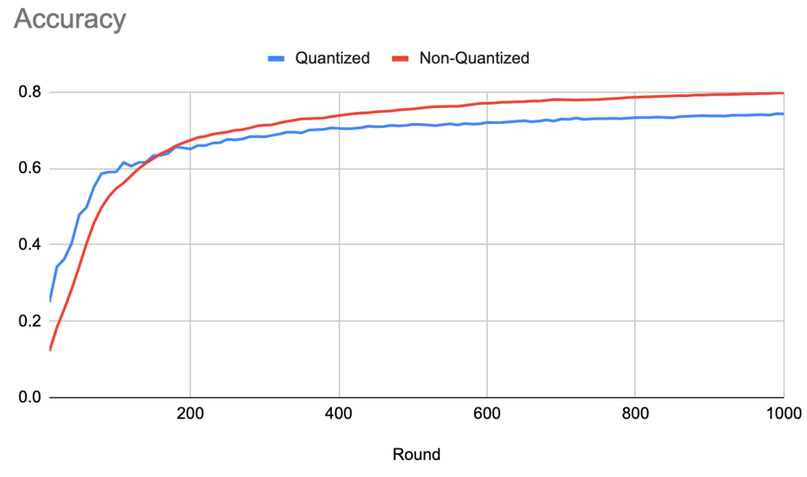
Receive parameter set θ_t;
Take a random sample L_t with probability $\frac{L}{N_i}$;
Choose subset $D_t = \{x_{j \in L_t}\}$ from D_i using Poisson subsampling function $\mathcal{S}^{po}(.)$;
Compute gradient $\mathbf{g}_t(D_t, \theta_t)$;
Clip gradient $\bar{\mathbf{g}}_t \leftarrow \mathbf{g}_t / max(1, \frac{\|\mathbf{g}_t\|_2}{S_q})$;
Create Gaussian dither vector $\vec{\mathbf{d}}$ by drawing $\frac{N_i}{L}$ i.i.d samples from $\mathcal{N}(0, \sigma^2 C^2 \mathbf{I})$ with $\sigma = N\Delta$;
Split $\bar{\mathbf{g}}_t$ into $\frac{N_i}{L}$ lattice vectors $\{\bar{\mathbf{g}}_{t1}, ..., \bar{\mathbf{g}}_{tL}\}$;
Randomize lattice vector $\bar{\mathbf{g}}_{tj} = \bar{\mathbf{g}}_{tj} + \vec{\mathbf{d}}(j)$;
Quantize each lattice vector $\bar{\mathbf{g}}_{tqj} = \Delta \texttt{round}\,(\bar{\mathbf{g}}_{tqj}/\Delta)$;
Encode the concatenated quantized gradient vector $\bar{\mathbf{g}}_{tq}$ using universal source coding method `bzip`;
Send encoded gradient vector $\bar{\mathbf{g}}_{e_{ti}}$ and codebook CB_{t_i} to PS;

Algorithm 2: Compressive (ϵ, δ)-DP Federated Learning - Parameter server (PS), Round t

Input: Encoded gradients of P participants $\{Bar\mathbf{g}_{e_{t1}}, ..., Bar\mathbf{g}_{e_{tP}}\}$ and codebooks $\{CB_{t_1}, ..., CB_{t_P}\}$
Parameters: PS learning rate η

Decode gradient vectors $\{Bar\mathbf{g}_{q_{t1}}, ..., Bar\mathbf{g}_{q_{tP}}\}$ using codebooks $\{CB_{t_1}, ..., CB_{t_P}\}$ for P clients;
Descend and calculate new model parameters $\theta_{t+1} = \theta_t - \eta \sum_{j=1}^{P} \bar{\mathbf{g}}_{q_{tj}}$;
Send updated model parameters θ_{t+1} to P clients for round t+1;

Results



- Results on EMNIST dataset. We simulate 25 participants. Data distribution is non-i.i.d, with each participant holding the data of 50 users. The charts are related to 8 quantization levels and vectors of length 5. For DP version of the experiment, we set $\delta = 0.0001$, $\varepsilon = 3.11$, and $\sigma = 0.25$. The compression ration after universal source coding via *bzip* is 8.03x.
- It should be noted that the results are preliminary. We are currently trying to enhance the noise shaping and the lattice design. More extensive experiments, comparing the proposed method with already available methods, e.g. DP-SGD, are also underway and will be published in a detailed manuscript along with the code.