Communication-Efficient Federated Learning via Dataset Distillation

Yanlin Zhou, George Pu, Xiyao Ma, Andy Xiaolin Li and Oliver Dapeng Wu

University of Florida and Cognization Lab

Problem

Existing federated learning algorithms suffer from:
- Expensive communication cost.
- Statistical heterogeneity, etc.

To achieve a performant model, tens of communication rounds are required to transmit unwieldy model weights under ideal circumstances and hundreds when data is poorly distributed.

Figure 2. Vanilla Dataset Distillation [3].

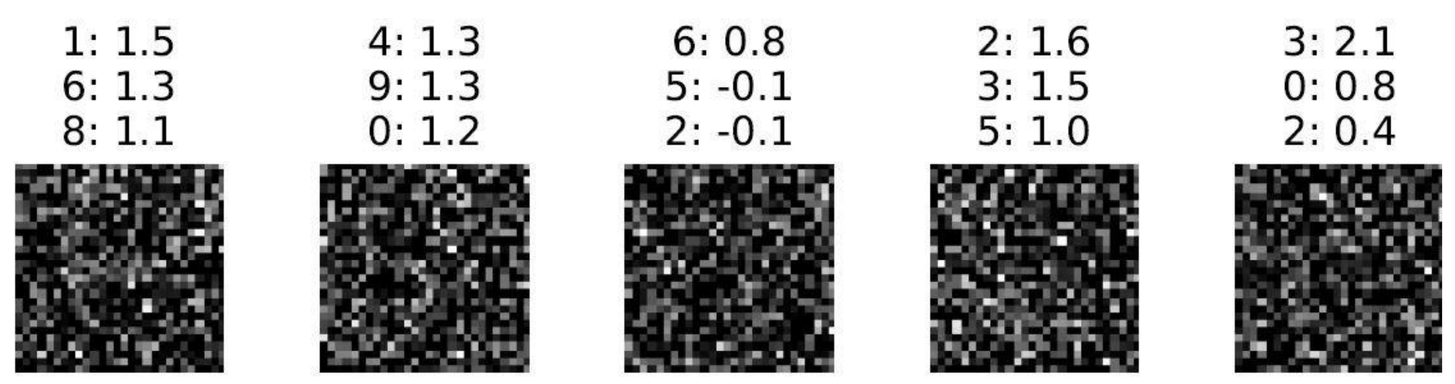

Figure 3. Distilled Images from MNIST.

> outset wed burroughs grossly contacted reginald anticipating dimitri returns nap housed feeds pitting woodward potts graduates attendant

Figure 3. Distilled sentences from IMDB sentiment analysis.

Impact

- DOSFL presents a brand-new paradigm of weight-less and gradient-less federated learning.
- The total communication cost of DOSFL is reduced by multiple orders of magnitude.
- DOSFL can be adopted with a hybrid system to jump-start a federated learning session.
- Distilled data targeted for specific models appear random to the human eye.

Table 1: Distillation Accuracy of DOSFL after One Round of Communication.

Dataset	Model	Additions	Baseline	Accuracy	Distillation Ratio
MNIST	LeNet	Soft Label	98.1%	95.6%	97.5%
IMDB	TextCNN	Soft Label, Soft Reset	86.1%	81.0%	94.1%
TREC-6	Bi-LSTM	Soft Label, Soft Reset	89.4%	83.60	93.5%

Future Work

We hope to:
1. Improve the non-IID performance, as well as incorporate existing FL advances.
2. Try different settings, i.e., imbalanced dataset, varying participation rate, malicious update, etc.
3. Apply DOSFL to regression tasks.

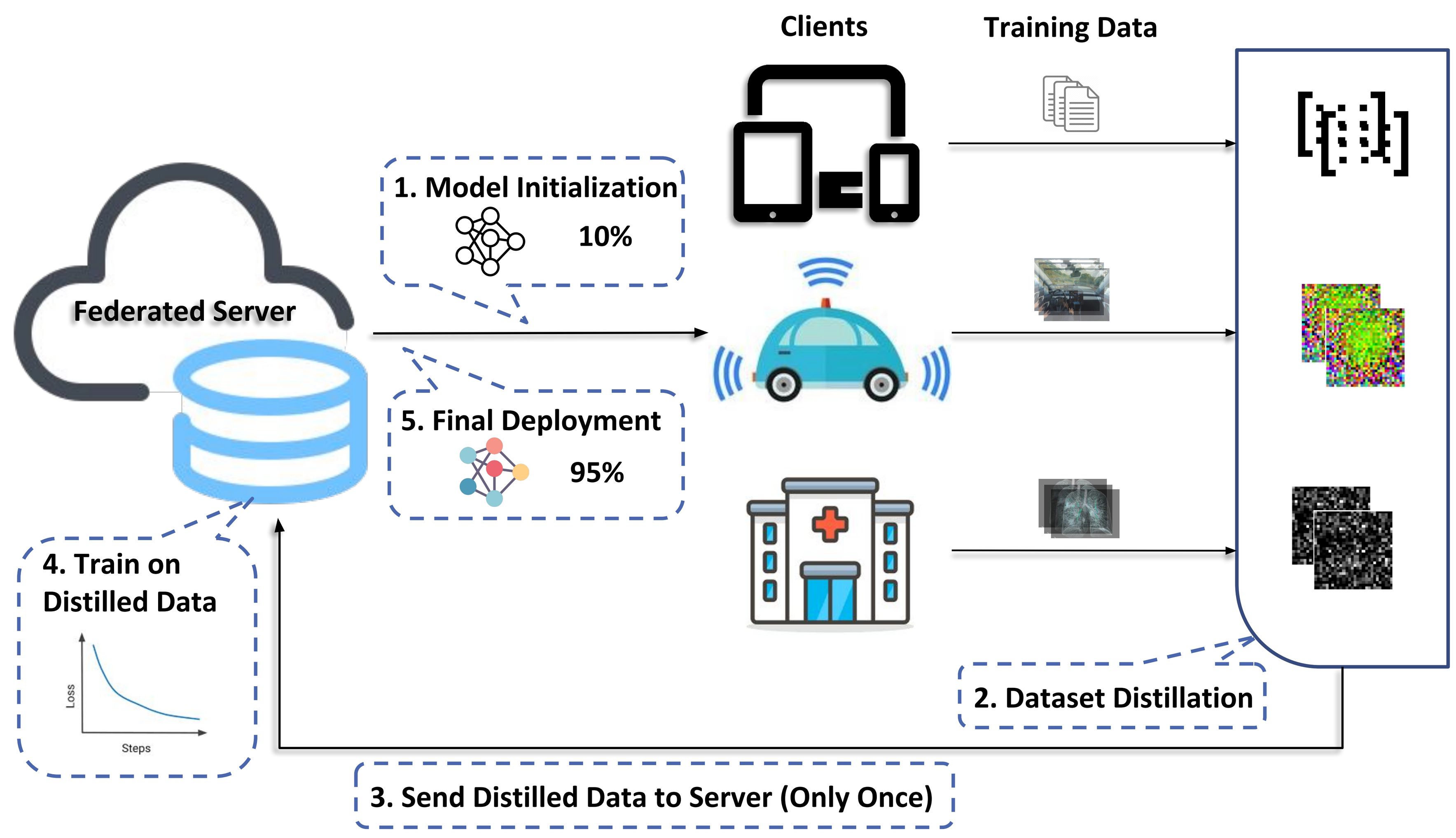

Figure 1. Distilled One-Shot Federated Learning.

Our Solution

We propose Distilled One-Shot Federated Learning (DOSFL) to reduce the communication round to only **ONE**.

We combat non-IID distribution issue by introducing:
- Soft label [1, 2]
- Soft reset: sample the starting parameters from a Gaussian distribution concentrated at the server's parameters
- Random masking: randomly selects a fraction of the distilled data each training iteration and replaces it with a random tensor

Algorithm 1 Distilled One-Shot Federated Learning

Initialize server weights θ_0
for clients $k = 1, \dots, N$ **do**
  DISTILLDATA(k, θ_0)
  Send distilled data to the server
end for
Merge distilled data into a single sequence $\{(\tilde{x}_j, \tilde{y}_j, \tilde{\eta}_j)\}_{j=1}^{NS_d}$
for $i = 0, 1, \dots, E_d - 1$ **do**
  for $j = 0, 1, \dots, NS_d$ **do**
    Number of adaptations $a = iNS_d + j$
    $\theta_{a+1} \leftarrow \theta_a - \tilde{\eta}_j \nabla \ell(\theta_a; \tilde{x}_j, \tilde{y}_j)$
  end for
end for

function DISTILLDATA(k, θ_0)
  Initialize $\{(\tilde{x}_j, \tilde{y}_j, \tilde{\eta}_j)\}_{j=1}^{S_d}$
  for $e = 1, 2, \dots, E$ **do**
    Get a minibatch $\mathcal{B}$ from the client's dataset $\mathcal{X}_k$
    for $i = 0, 1, \dots, E_d - 1$ **do**
      for $j = 0, 1, \dots, S_d - 1$ **do**
        Number of adaptations $a = iS_d + j$
        $\theta_{a+1} = \theta_a - \tilde{\eta}_j \nabla \ell(\theta_a; \tilde{x}_j, \tilde{y}_j)$
      end for
    end for
    $\tilde{x} \leftarrow \tilde{x} - \alpha \nabla_{\tilde{x}} \ell(\theta_{E_d S_d}; \mathcal{B})$
    $\tilde{\eta} \leftarrow \tilde{\eta} - \alpha \nabla_{\tilde{\eta}} \ell(\theta_{E_d S_d}; \mathcal{B})$
    if *soft_label* **then**
      $\tilde{y} \leftarrow \tilde{y} - \alpha \nabla_{\tilde{y}} \ell(\theta_{E_d S_d}; \mathcal{B})$
    end if
  end for
  return $\{(\tilde{x}_i, \tilde{y}_i, \tilde{\eta}_i)\}_{i=1}^{S_d}$
end function

Table 2: Communication comparison between DOSFL and FedAvg.

Dataset	Model	n_{data}	B_d	Θ	Break even round	Comm. reduction
MNIST	LeNet	28×28	10	61,706	38.65	53.4%
IMDB	TextCNN	200×100	1	120,601	8.29	66.8%
TREC-6	Bi-LSTM	30×100	1	404,406	0.14	99.2%

References

1. Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network.arXiv preprint arXiv:1503.02531, 2015.
2. Ilia Sucholutsky and Matthias Schonlau. Soft-label dataset distillation and text dataset distillation, 2019
3. Tongzhou Wang, Jun-Yan Zhu, Antonio Torralba, and Alexei A Efros. Dataset distillation.arXiv preprint arXiv:1811.10959, 2018.

UNIVERSITY of FLORIDA

NSF Center for Big Learning