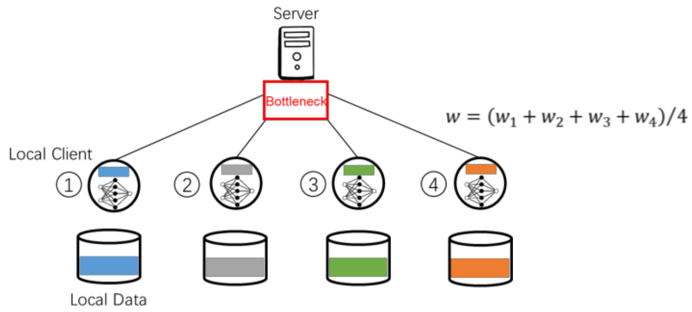
DYNAMIC DECENTRALIZED FEDERATED LEARNING

Shenghong Dai, Kangwook Lee, Suman Banerjee @ UW Madison

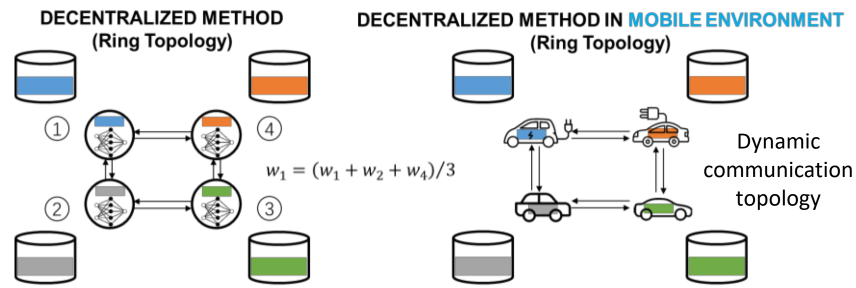
FEDERATED LEARNING & ITS LIMITATIONS

- Federated Learning (FL) trains a shared model through training decentralized data over clients while communicating only model updates
- The server may become a bottleneck as the number of clients increases



DYNAMIC DECENTRALIZED FL

- Decentralized FL: clients share their model updates with their neighbors instead of the central coordinator



PROPOSED ALGORITHM

- Dynamic decentralized FL on the ith client
- **Input:** initial point $w_{0,i}$, communication matrix E, the number of iterations K
  - **for** $k = 0, \dots, K-1$ **do**

    Random sample data; Update the local model using the sampled data;

    Average the local model with neighbors: $w_{k+1,i} = \sum_{j=1}^{n} E_{ij} w_{k,j}$

- where every entry E_{ij} of the communication matrix $E \in \mathbb{R}^{n \times n}$ shows how much will client j have impact on client i

$$E_{ij} = \begin{cases} 1/n & \text{if } j \neq i \text{ and } j \text{ is a neighbor of } i, \\ 1 - |\text{neighbor of } i|/n & \text{if } j = i, \\ 0 & \text{otherwise.} \end{cases}$$

CONVERGENCE RATE ANALYSIS

- We consider solving the following optimization problem

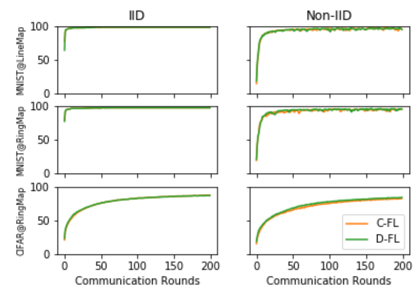
$$\min_{w \in \mathbb{R}^N} \frac{1}{n} \sum_{i=1}^{n} f_i(w)$$

where each $f_i: \mathbb{R}^N \to \mathbb{R}$ is the local objective function of client i

- Assumptions: Lipschitz gradient & Spectral gap & Bounded variance & Start from 0 & the clients are connected sufficiently often in time
- Theorem: the convergence rate for dynamic decentralized FL is $O(\frac{1}{K} + \frac{1}{\sqrt{nK}})$ if the number of iterations K is large enough

EXPERIMENTAL RESULTS

Test Accuracy in various settings. Results validate our theory since D-FL has the same convergence rate.



FUTURE WORK

- Generate a dataset from virtual agents in a virtual world
- Compared with the pre-existing FL datasets, this dataset is
  - more conform to the actual data collected from mobile nodes
  - naturally partitioned
  - having a realistic modeling for the dynamic graph

MLWiNS: Distributed Learning for the Nomadic Edge (CNS-2003129); PI: Suman Banerjee

References

Lian et al. (2017). Can decentralized algorithms outperform centralized algorithms?
Nedic et al. (2008). Distributed subgradient methods and quantization effects.