July 2022

TEXAS
The University of Texas at Austin

ARCHITECTURE AGNOSTIC FEDERATED LEARNING FOR NEURAL NETWORKS

International Conference of Machine Learning 2022

DISHA MAKHIJA, XING HAN, NHAT HO, JOYDEEP GHOSH

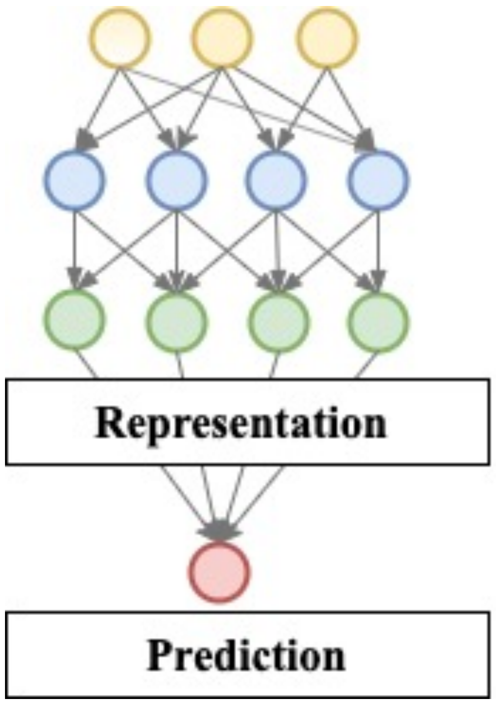
Outline

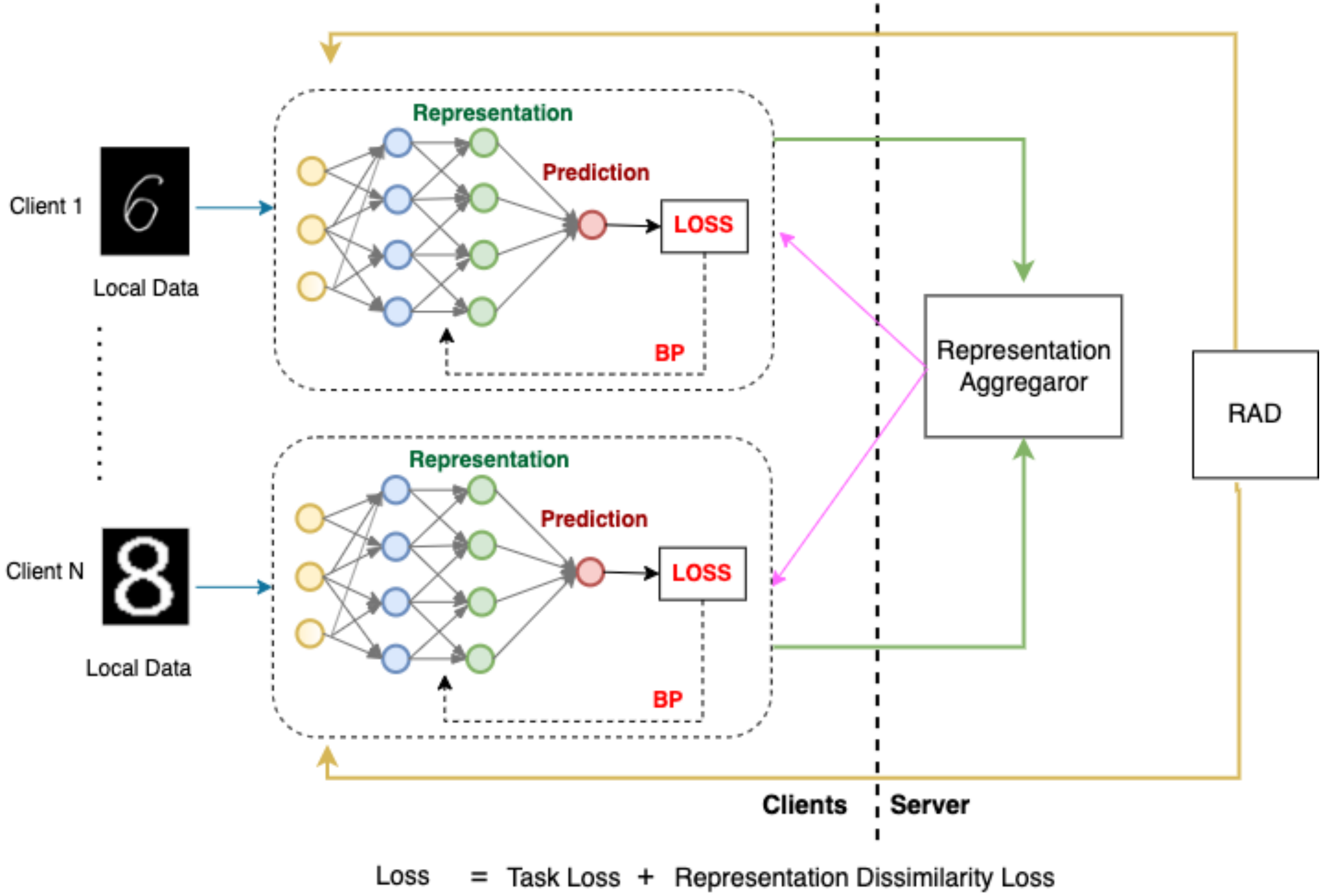
- Motivation
- Proposed Approach
- Experimental Results

Motivation

- Standard FL algorithms assume each client use the same architecture
  - But clients may differ in their compute resources and local data size/properties
  - May have pre-existing (uncoordinated) solutions
  - Identical architecture requirement is severe, even impractical
- Aggregation of client models to generate single global model is not straightforward due to facts like permutation invariance of neurons, non-linearities etc.
- We propose a systematic framework for architecture-agnostic federated learning called FedHeNN

Proposed Approach : FedHeNN





Proposed Approach : FedHeNN

- FedHeNN for Homogeneous Clients

$$\min_{\mathcal{W}_i} \mathcal{L}_i = \mathcal{F}(\mathcal{W}_i) + \eta\, \mathrm{d}(\Phi_i(\mathbf{X}; \mathcal{W}_i), \Phi_{\text{global}}(\mathbf{X}; \hat{\mathcal{W}}(t-1))).$$

- FedHeNN for Heterogeneous Clients

$$\min_{\mathcal{W}_i} \mathcal{L}_i = \mathcal{F}(\mathcal{W}_i) + \eta\, \mathrm{d}(\Phi_i(\mathbf{X}; \mathcal{W}_i), \bar{\Phi}(\mathbf{X}; (t-1))).$$

where $\bar{\Phi}(\mathbf{X}; (t-1)) = \sum_{j=1}^{N} w_j \Phi_j(\mathbf{X}; \mathcal{W}_j(t-1))$

Distance Function

- Use a kernel-based distance metric to allow comparison between NNs of different widths
- Specifically, we use CKA(centered kernel alignment) whose linear version is given by

$$\text{Linear CKA}(K_i, K_j) =$$

$$\text{Linear CKA}(A_i A_i^T, A_j A_j^T) = \frac{||A_j^T A_i||_F^2}{||A_i^T A_i||_F ||A_j^T A_j||_F}.$$

Experimental Results

DATA SET(SETTING)	FEDAVG	FEDPROX	FEDHENN GLOBAL
CIFAR-10(100 CLIENTS, 2 CLS/CLIENT)	44.29 ± 0.5	53.8 ± 2.3	$\mathbf{68.8 \pm 2.1}$
CIFAR-10(100 CLIENTS, 5 CLS/CLIENT)	58.14 ± 0.7	63.3 ± 2.0	$\mathbf{70.19 \pm 2.0}$
CIFAR-10(500 CLIENTS, 2 CLS/CLIENT)	42.7 ± 0.4	50.46 ± 1.4	$\mathbf{65.4 \pm 0.8}$
CIFAR-10(500 CLIENTS, 5 CLS/CLIENT)	56.8 ± 0.5	55.2 ± 1.2	$\mathbf{64.7 \pm 0.7}$
CIFAR-100(100 CLIENTS, 20 CLS/CLIENT)	28.6 ± 0.8	27.3 ± 1.1	$\mathbf{44.2 \pm 0.7}$
SENTIMENT140(100 CLIENTS, 2 CLS/CLIENT)	52.6 ± 0.4	52.7 ± 1.0	$\mathbf{52.7 \pm 0.01}$

Table 1: Average test accuracy of FedHeNN computed for the common global model as compared to the baselines with global models.

DATA SET(SETTING)	FEDREP	FEDHENN HOMO	FEDHENN HETERO
CIFAR-10(100 CLIENTS, 2 CLS/CLIENT)	85.7 ± 0.4	94.7 ± 1.1	88.9 ± 0.35
CIFAR-10(100 CLIENTS, 5 CLS/CLIENT)	72.4 ± 1.2	84.37 ± 1.5	73.01 ± 0.3
CIFAR-10(500 CLIENTS, 2 CLS/CLIENT)	78.9 ± 0.6	86.5 ± 0.9	82.02 ± 0.8
CIFAR-10(500 CLIENTS, 5 CLS/CLIENT)	58.14 ± 0.21	73.32 ± 1.23	61.74 ± 0.6
CIFAR-100(100 CLIENTS, 20 CLS/CLIENT)	38.85 ± 0.9	62.89 ± 0.8	43.36 ± 0.2
SENTIMENT140(100 CLIENTS, 2 CLS/CLIENT)	69.8 ± 0.4	72.6 ± 0.3	71.5 ± 0.5

Table 2: Average test accuracy of FedHeNN computed for the personalised models as compared to the baselines with personalised models.

Thank you!