Bridging Generalization Gap of Heterogeneous Federated Clients Using Generative Models

Ziru Niu[1], Hai Dong[1], A. K. Qin [2]

ziru.niu@student.rmit.edu.au, hai.dong@rmit.edu.au, kqin@swin.edu.au

[1]RMIT University, Melbourne, VIC, Australia
[2]Swinburne University of Technology, Hawthorn, VIC, Australia

Abstract

In this paper, we propose a model-heterogeneous FL framework that enhances clients' generalization performance on unseen data without relying on parameter aggregation. Instead of model parameters, clients share feature distribution statistics (mean and covariance) with the server. Then each client trains a variational transposed convolutional neural network using Gaussian latent variables sampled from these distributions, and use it to generate synthetic data. By fine-tuning local models with the synthetic data, clients achieve significant improvement of generalization ability.

Background

The non-iid data of clients has become a widespread problem in federated learning (FL). Firstly, clients comprise edge devices distributed across various geographical locations, which naturally collect non-identically-independent data. Secondly, single clients usually collect real-time data with frequent distribution shifts in practical scenarios, such as weather forecasting.

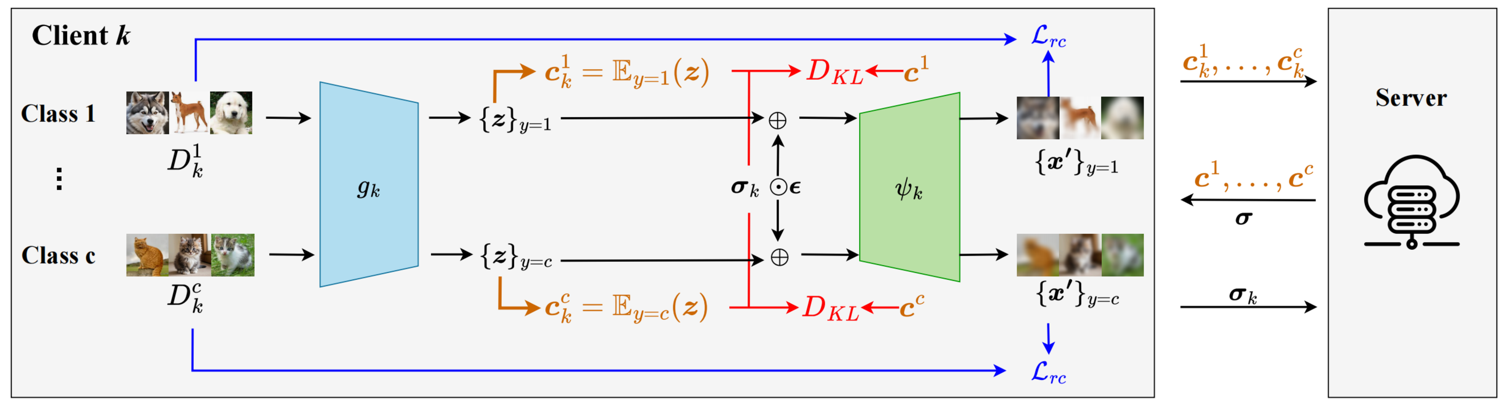
To overcome both inter-client and intra-client non-iid, improving the *generalization* ability (i.e., the ability to predict unseen data correctly) of clients has become a vital matter. Most state-of-the-art works are developed based on the assumption that all clients share the same model architecture. In practice, clients are likely to obtain personalized models with diverse architectures, making model-ghomogeneous methods prohibitive, which necessitates alternative strategies for enhancing the generalization of heterogeneous local models.

Methodology

We propose **FedVTC** (Federated Learning with Varational Transposed Convolution)

FedVTC Algorithm

1. Clients train a shared VTC model ψ with a universal architecture.
2. Clients share prototypes $\boldsymbol{c}$, SD $\boldsymbol{\sigma}$ with the server in each round.
3. Clients generate synthetic samples using $\boldsymbol{c}, \boldsymbol{\sigma}$ and the VTC model ψ.
4. Clients fine-tune local models using the synthetic samples to boost generalization.



VTC Training

A VTC model ψ is trained using loss function $\mathcal{L}_{tc}$, where:

$$\mathcal{L}_{tc} = \mathcal{L}_e + \lambda \mathcal{L}_{dm}$$

$\mathcal{L}_e = -$ELBO (Evidence Lower Bound):

$$\text{ELBO} = log\, p_{\psi_k}(\boldsymbol{x}'|\boldsymbol{z}, \boldsymbol{\sigma_k}) - D_{KL}(q_{g_k}(\boldsymbol{z}|\boldsymbol{x})||p(\boldsymbol{z}))$$
$$\approx \|\boldsymbol{x'} - \boldsymbol{x}\|_2^2 + \frac{1}{2}\left[(\boldsymbol{z} - \boldsymbol{c}^y)^T(\boldsymbol{z} - \boldsymbol{c}^y) + tr(\Sigma) - p - log|\Sigma|\right]$$

Σ is a diagnoal matrix with $\Sigma_{ii} = \boldsymbol{\sigma}_i$.
$\mathcal{L}_{dm}$ is the distribution-matching loss:

$$\mathcal{L}_{dm} = \|g_k(\psi_k(\boldsymbol{v})) - \boldsymbol{c}^y\|_2^2$$

	Dir	FedProto	FedTGP	FedGen	CCVR	FedType	pFedAFM	**FedVTC**
MNIST	0.1	83.4±0.4	85.4±0.5	81.1±1.0	85.5±0.4	83.9±0.3	85.6 ± 0.5	**88.7**±0.7
	1.0	86.1±0.3	86.8±0.4	81.9±0.3	86.2±0.3	86.7±0.5	87.2 ± 0.4	**90.1**±0.1
CIFAR10	0.1	39.3±0.7	40.9±0.2	39.1±0.2	39.3±0.2	39.7±0.3	38.9 ± 0.9	**46.9**±0.6
	1.0	41.1±0.2	42.1±0.7	39.7±0.5	41.0±0.1	40.2±0.6	40.5 ± 0.5	**51.7**±0.7
CIFAR100	0.1	26.2±0.2	29.2±0.3	27.3±0.8	27.6±0.9	31.3±1.5	32.2 ± 1.0	**36.3**±0.8
	1.0	30.4±0.2	31.8±0.5	30.8±0.3	29.3±0.4	34.4±0.6	34.9 ± 0.6	**40.4**±0.3
Tiny-ImageNet	0.1	23.7±0.3	24.9±0.8	23.6±0.1	24.1±0.3	24.8±0.4	24.2 ± 0.8	**30.2**±0.3
	1.0	26.4±0.2	27.2±0.3	26.1±0.1	26.6±0.8	31.6±0.4	27.3 ± 0.2	**35.8**±0.5

Table 1: The average generalization accuracy (in %) of each client's local model on the global validation dataset (with mean ± SD).

	FedProto	FedTGP	FedGen	CCVR	FedType	pFedAFM	**FedVTC**
MNIST	0.94	0.94	13.15	46.61	12.37	1.94	**0.70**
CIFAR10	3.93	3.93	54.58	807.46	37.45	5.88	**2.90**
CIFAR100	39.32	39.32	91.07	825.16	72.85	41.27	**26.49**
Tiny-ImageNet	78.64	78.64	844.82	923.46	179.61	147.10	**52.71**

Table 2: The total volume of data transmission (in GB).

Evaluation

FedVTC achieves higher generalization accuracy and lower communication costs as shown in Table 1 and Table 2.

Synthetic Samples

A fraction of the synthetic samples on MNIST and CIFAR10 are shown below: