Communication Efficient Training of Federated Model Over Unbalanced Labels

Yawei Zhao, Qinghe Liu, Mingming Jiang, Kunlun He

***Abstract*—xxx**

***Index Terms*—xxxx**

I. Introduction

xxx

II. Formulation

A. Personalized Representation based on Similarity Network

Personalized models are built based on the similarity network. The similarity network measures the similarity of data distribution under data/feature/model space.

- **Data space.** In the case, local datasets of every node are used to construct a *kernel* matrix. The similarity of data distribution is measured by xxxxx.
- **Feature space.** In the case, local datasets of every node are used to construct a *covariance* matrix. It represents the dependence structure among features. The similarity of data distribution is measured based on the distance between covariance matrices.
- **Model space.** In the case, local model of every node is trained by using the local dataset. The similarity of data distribution is measured based on the distance between local models.

Based on similarity under those space, the similarity network $\mathcal{G} = \{\mathcal{N}, \mathcal{E}\}$ is built by using the KNN method [**?**]. $\mathcal{N} = \{1, 2, ..., N\}$ represents the node set, and $\mathcal{E} = \{e_{i,j} : i \in \mathcal{N}, j \text{ is the node } i\text{'s neighbour}\}$.

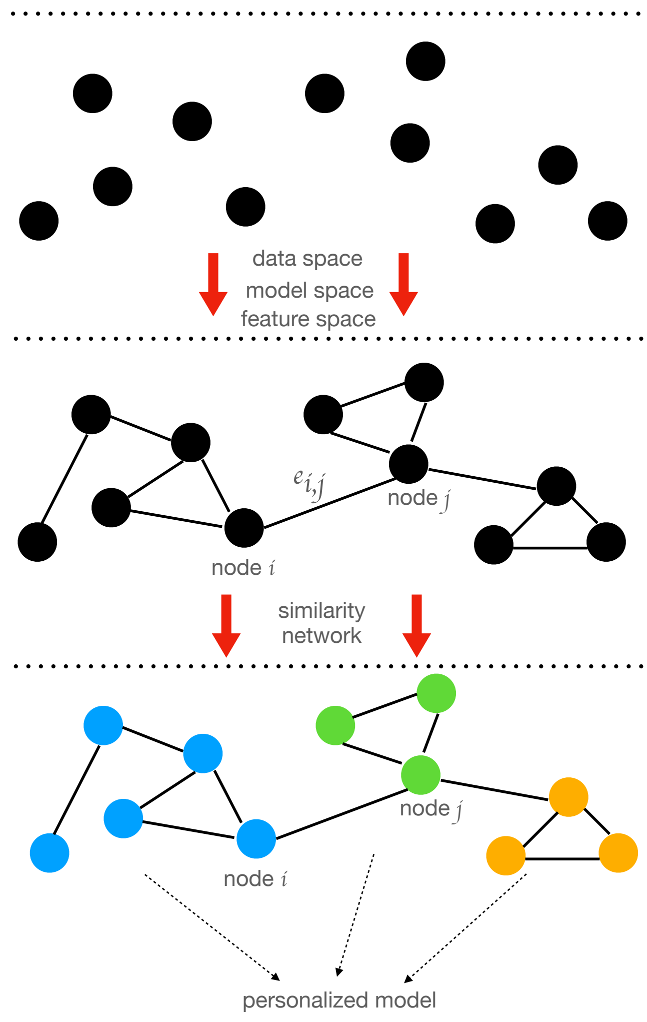

Fig. 1. Personalized representation based on similarity network

B. Final Formulation

Given the mapping matrix $\mathbf{M} \in \mathbb{R}^{d \times d_1}$ and $\mathbf{N} \in \mathbb{R}^{d \times d_2}$, the objective problem is formulated by

$$\min_{\{\mathbf{x}^{(n)}\}_{n=1}^{N}} \sum_{n \in \mathcal{N}} f_n\left(\mathbf{x}^{(n)}; \mathbf{A}_n\right) + \lambda \sum_{\substack{e_{i,j} \in \mathcal{E}, \\ \forall i,j \in \mathcal{N}}} \left\| \mathbf{z}^{(i)} - \mathbf{z}^{(j)} \right\|_p, \quad (1)$$

subject to:

$$\mathbf{x}^{(n)} = \mathbf{M}\mathbf{x} + \mathbf{N}\mathbf{z}^{(n)}, \qquad \forall n \in \mathcal{N}, \mathbf{x} \in \mathbb{R}^{d_1}, \mathbf{z}^{(n)} \in \mathbb{R}^{d_2}. \quad (2)$$

Here, $p \in \{1, 2, \infty\}$. $\mathbf{M}$ and $\mathbf{N}$ has special structure, where every row of them has at most one non-zero value, and the non-zero value is 1.

C. Federated Optimization Methods

The update of local model $\mathbf{x}_{t+1}^{(n)}$ is obtained by solving the following problem.

$$\min_{\mathbf{x}^{(n)} \in \mathbb{R}^d} f_n\left(\mathbf{x}^{(n)}; \mathbf{A}_n\right) + \lambda \sum_{\substack{e_{n,j} \in \mathcal{E}, \\ \forall j \in \mathcal{N}}} \left\| \mathbf{z}^{(n)} - \mathbf{z}^{(j)} \right\|, \quad (3)$$

subject to:

$$\mathbf{x}^{(n)} = \mathbf{M}\mathbf{x} + \mathbf{N}\mathbf{z}^{(n)}. \quad (4)$$

Yawei Zhao, Qinghe Liu, Mingming Jiang, and Kunlun He are with the Medical Big Data Research Center, Chinese PLA General Hospital, Beijing, 100039, China. E-mail: csyawei.zhao@gmail.com, Liuqinghe9638@163.com, jiangmingming1994@163.com, kunlunhe@plagh.org.

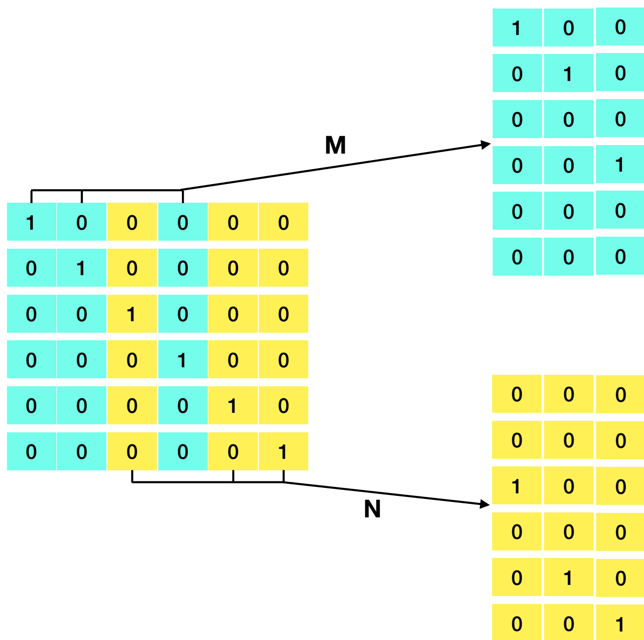

Fig. 2. for example: $\mathbf{M}$ and $\mathbf{N}$

By using SGD [**?**], the local model is updated by performing the following problem.

$$\min_{\mathbf{x}^{(n)} \in \mathbb{R}^d} \left\langle \mathbf{g}_t^{(n)}, \mathbf{x}^{(n)} - \mathbf{x}_t \right\rangle + \lambda \sum_{\substack{e_{n,j} \in \mathcal{E}, \\ \forall j \in \mathcal{N}}} \left\| \mathbf{z}^{(n)} - \mathbf{z}^{(j)} \right\| + \frac{1}{2\eta} \left\| \mathbf{x}^{(n)} - \mathbf{x}_t \right\|^2, \tag{5}$$

subject to:

$$\mathbf{x}^{(n)} = \mathbf{M}\mathbf{x} + \mathbf{N}\mathbf{z}^{(n)}. \tag{6}$$

On client.

On server.

III. Communication Efficient Training

IV. Empirical Studies

Acknowledgment

This work was supported by the xx.