

# 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 [?].  $\mathcal{N} = \{1, 2, \dots, N\}$  represents the node set, consisting of  $N$  nodes.  $\mathcal{E} = \{e_{i,j} : i \in \mathcal{N}, j \text{ is the node } i\text{'s neighbour}\}$  represents the edge set, consisting of  $M$  edges.

### 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

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

subject to:

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

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.

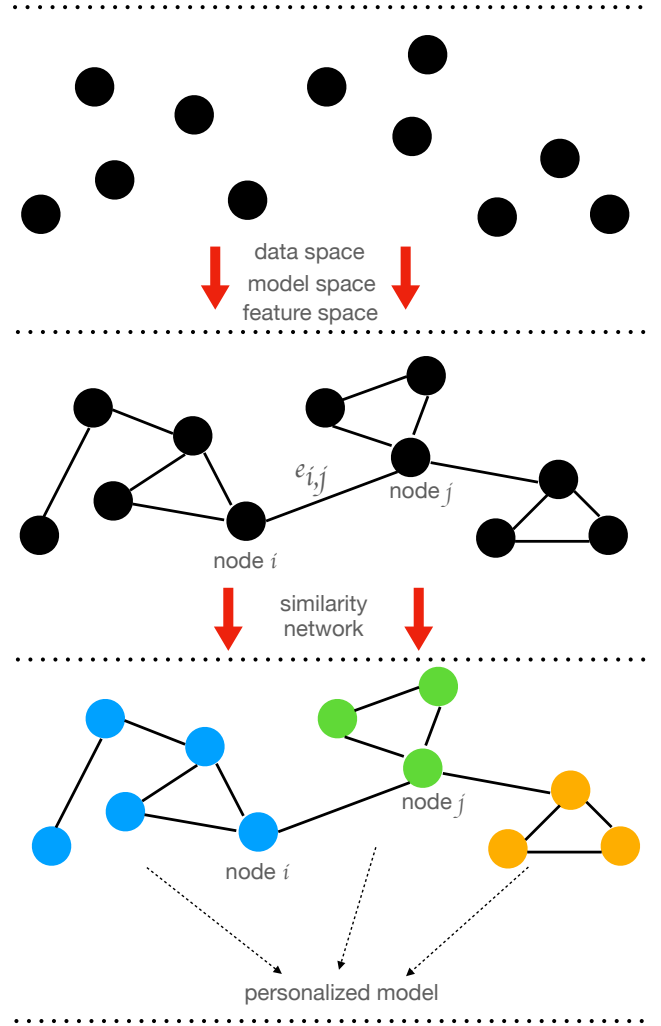


Fig. 1. Personalized representation based on similarity network

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.

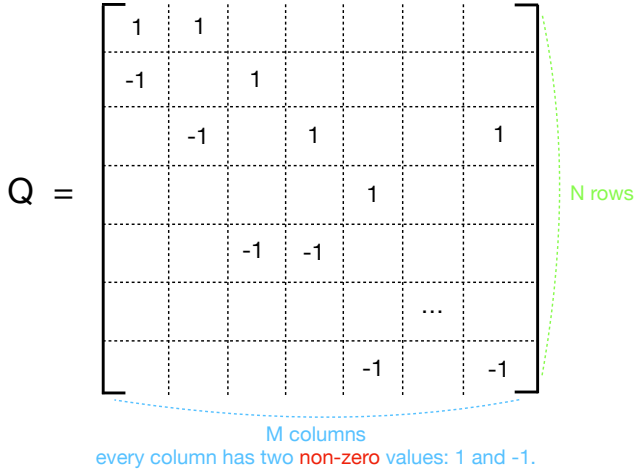
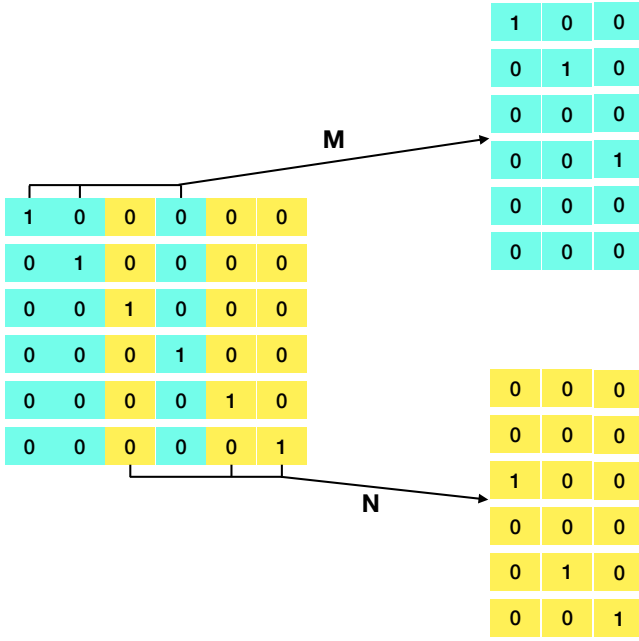
The formulation can be equally transformed as follows.

$$\min_{\{\mathbf{x}^{(n)}\}_{n=1}^N} \frac{1}{N} \sum_{n=1}^N f_n(\mathbf{x}^{(n)}; \mathbf{D}_n) + \lambda \|\mathbf{Z}\mathbf{Q}\|_{1,p}, \quad (3)$$

subject to:

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

Here,  $\mathbf{Z} \in \mathbb{R}^{d_2 \times N}$ ,  $\mathbf{Q} \in \mathbb{R}^{N \times M}$ .  $N$  and  $M$  represent the total number of nodes and edges in the network  $\mathcal{G}$ , respectively.  $\mathbf{Z}$

Fig. 2.  $\mathbf{Q}$  matrix.Fig. 3. for example:  $\mathbf{M}$  and  $\mathbf{N}$ 

represents some a variable matrix, consisting of  $N$  variables as columns, that is,  $\mathbf{Z} = [\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \dots, \mathbf{z}^{(N)}]$ .  $\mathbf{Q}$  is the given auxiliary matrix, which has  $M$  columns and every column has two non-zero values: 1 and  $-1$ . Note that  $\|\cdot\|_{1,p}$  is denoted by  $\ell_{1,p}$  norm. Given a matrix  $\mathbf{U} \in \mathbb{R}^{d_2 \times M}$ , it is defined by

$$\|\mathbf{U}\|_{1,p} := \sum_{m=1}^M \|\mathbf{U}_{:,m}\|_p.$$

### C. Optimization via Alternative method

$$\min_{\{\mathbf{x}^{(n)}\}_{n=1}^N} \frac{1}{N} \sum_{n=1}^N f_n(\mathbf{x}^{(n)}; \mathbf{D}_n) + \lambda \|\mathbf{ZQ}\|_{1,p}, \quad (5)$$

subject to:

$$\mathbf{x}^{(n)} = \mathbf{Mx} + \mathbf{Nz}^{(n)}. \quad (6)$$

The variable  $\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}\}$  is obtained by alternatively optimizing  $\mathbf{x}$  and  $\{\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \dots, \mathbf{z}^{(N)}\}$ .

**Optimizing  $\mathbf{x}$  by given  $\{\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \dots, \mathbf{z}^{(N)}\}$ .**  $\mathbf{x}$  is optimized by solving the following problem.

$$\min_{\mathbf{x} \in \mathbb{R}^{d_1}} \frac{1}{N} \sum_{n=1}^N f_n(\mathbf{Mx} + \mathbf{Nz}^{(n)}; \mathbf{D}_n). \quad (7)$$

By using the data-driven stochastic optimization method [?], we need to perform the following problem to obtain  $\mathbf{x}_{t+1}$ , iteratively.

$$\begin{aligned} & \mathbf{x}_{t+1} \\ &= \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^{d_1}} \frac{1}{N} \sum_{n=1}^N \left\langle \mathbf{M}^{-1} \mathbf{g}_t^{(n)}, \mathbf{x} \right\rangle + \frac{1}{2\eta} \|\mathbf{x} - \mathbf{x}_t\|^2, \end{aligned}$$

where  $\mathbf{g}_t^{(n)}$  is a stochastic gradient of  $f_n$  with  $\mathbf{Mx}_t + \mathbf{Nz}_t^{(n)}$  by using data drawn from the local dataset  $\mathbf{D}_n$ .

**Optimizing  $\{\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \dots, \mathbf{z}^{(N)}\}$  by given  $\mathbf{x}$ .**

$$\min_{\{\mathbf{z}^{(n)}\}_{n=1}^N} \frac{1}{N} \sum_{n=1}^N f_n(\mathbf{Mx} + \mathbf{Nz}^{(n)}; \mathbf{D}_n) + \lambda \|\mathbf{ZQ}\|_{1,p}. \quad (8)$$

By using the data-driven stochastic optimization method [?], we need to perform the following problem.

$$\min_{\{\mathbf{z}^{(n)}\}_{n=1}^N} \frac{1}{N} \sum_{n=1}^N \left\langle \mathbf{N}^{-1} \mathbf{g}_t^{(n)}, \mathbf{z}^{(n)} \right\rangle + \lambda \|\mathbf{ZQ}\|_{1,p} + \frac{1}{2\eta} \|\mathbf{Z} - \mathbf{Z}_t\|_F^2. \quad (9)$$

$\mathbf{g}_t^{(n)}$  is a stochastic gradient of  $f_n$  with  $\mathbf{Mx}_t + \mathbf{Nz}_t^{(n)}$  by using stochastic data drawn from the local dataset  $\mathbf{D}_n$ .

**On client.**

**On server.**

## III. COMMUNICATION EFFICIENT TRAINING

### IV. EMPIRICAL STUDIES

#### ACKNOWLEDGMENT

This work was supported by the xx.