Privacy-Preserving Asynchronous Vertical Federated Learning Algorithms for Multiparty Collaborative Learning

Vertically Partitioned data

- Definition :
  - the data locate at multiple (two or more) data holders, and each maintains its own records of different feature sets with common entities, which are called vertically partitioned (VP) data.
  - Example: a digital finance company, an E-commerce company, and a bank collect different information about the same person.

- Need of efficient FL algorithms on the VP data
  - If the person submits a loan application to the digital finance company, it might want to evaluate the credit risk of approving this financial loan by comprehensively utilizing the information stored in all three parties.

Already Existing Work

Synchronous version of algorithms :

- SGD
- SVRG (stochastic variance reduced gradient)
- SAGA (stochastic average gradient)

Their Contributuion

Asynchronous version of algorithms :

- AFSGD-VP
- AFSVRG-VP (stochastic variance reduced gradient)
- AFSAGA-VP (stochastic average gradient)

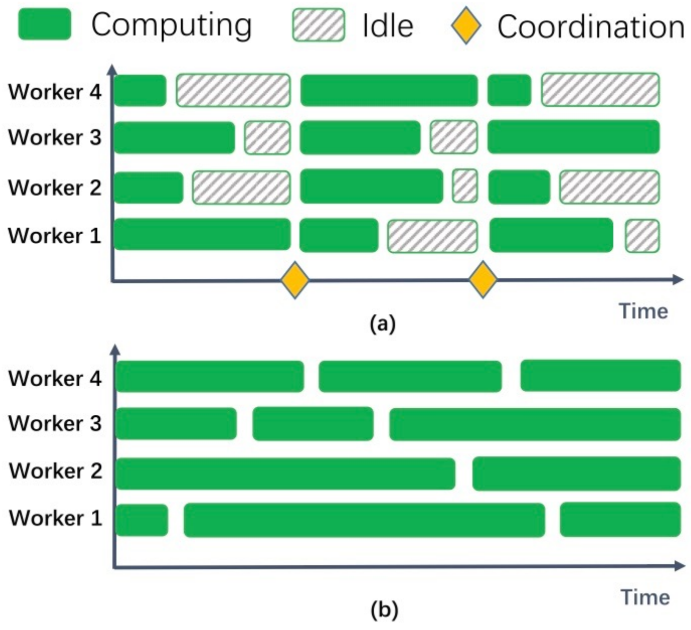

Fig. 1. (a) Synchronous computation versus (b) asynchronous computation.

synchronous computation is much more inefficient than the asynchronous computation because it wastes a lot of computing resources to be idle

Problem Statement

In this article, we consider the model in a linear form of $w^T x$. Given a training set $\mathcal{S} = \{(x_i, y_i)\}_{i=1}^{l}$, where $x_i \in \mathbb{R}^d$ and $y_i \in \{+1, -1\}$ for binary classification or $y_i \in \mathbb{R}$ for regression,[1] the loss function with respect to the sample (x_i, y_i) and the model weights w can be formulated as $L(w^T x_i, y_i)$. Thus, we consider to optimize the following regularized empirical risk minimization problem:

$$\min_{w \in \mathbb{R}^d} f(w) = \frac{1}{l} \sum_{i=1}^{l} \underbrace{L(w^T x_i, y_i) + g(w)}_{f_i(w)} \tag{1}$$

where $g(w)$ is a regularization term, and each $f_i : \mathbb{R}^d \to \mathbb{R}$

How data is arranged ?

learning applications, the input of training sample (x, y) is partitioned vertically into q parts, i.e., we have a partition $\{\mathcal{G}_1, \ldots, \mathcal{G}_q\}$ of d features. Thus, we have $x = [x_{\mathcal{G}_1}, x_{\mathcal{G}_2}, \ldots, x_{\mathcal{G}_q}]$, where $x_{\mathcal{G}_\ell} \in \mathbb{R}^{d_\ell}$ is stored on the ℓth worker, and $\sum_{\ell=1}^{q} d_\ell = d$. According to whether the label is included in a worker, we divide the workers into two types: one is the active worker and the other is passive worker, where the active worker is the data provider who holds the label of a sample beside the partial input of a sample, and the passive worker only has the partial input of a sample without label information. The active worker would be a dominating server in federated learning, while passive workers play the role of clients [13]. We let D^ℓ denote the data stored on the ℓth worker. Note that the labels y_i are distributed to active

Goal

Goal: *Make active workers cooperate with passive workers to solve the regularized empirical risk minimization problem (1) on the VP data $\{D^\ell\}_{\ell=1}^{q}$ in parallel and asynchronously with the SGD and its SVRG and SAGA variants while keeping the VP data private.*

System Structure of this algorithm

- *Tree-Structured Communication:*
- *Data and Model Privacy:*

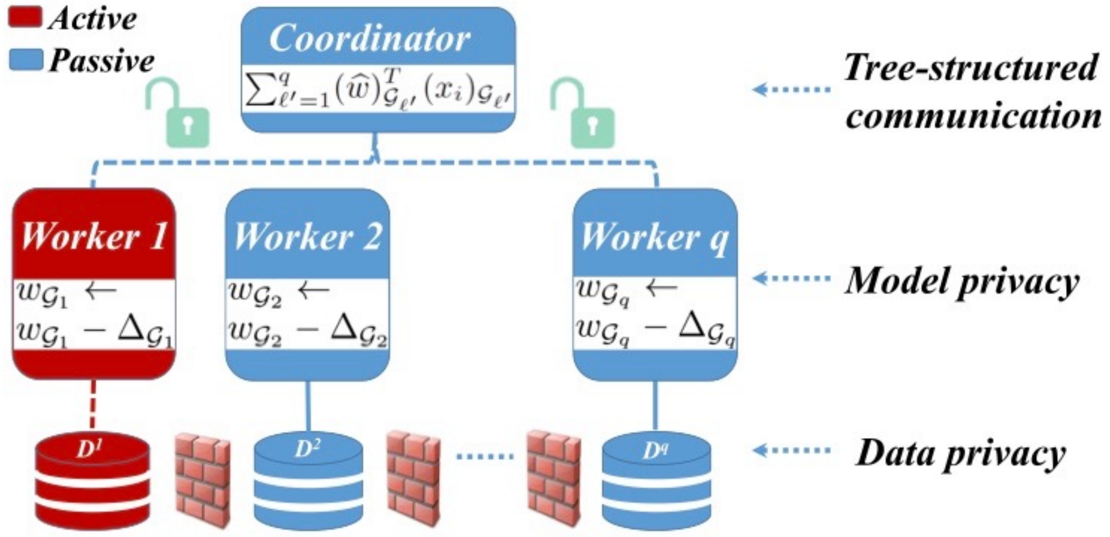

Fig. 2. System structure of our privacy-preserving asynchronous federated learning algorithms.