Abstract

We study two VFL paradigms—FedCVT (semi-supervised cross-view training) and communication-efficient VFL for limited overlap (Sun et al.). Addressing sparse sample alignment and bandwidth constraints, we analyze goals, assumptions, and architectures. A minimal two-client simulation illustrates core design trade-offs and how unlabeled data can be exploited in one-/few-shot settings. We conclude with practical guidance for low-resource deployments.

Distributed Machine Learning

Understanding & Comparing Modern VRL Technics

|  |  |  |
| --- | --- | --- |
| **NAME** | **BITS ID** | **CONTRIBUTION** |
| SUBHRANSU MISHRA | 2023AC05489 | 100% |
| DULAL DAS | 2023AC05041 | 100% |
| LAKSHMISRINIVAS PERAKAM | 2023AC05540 | 100% |
| ARCHAN GHOSH | 2023AC05402 | 100% |

## Part A

Introduction

This analysis explores the underlying principles and architectural differences between two novel Vertical Federated Learning (VFL) approaches: FedCVT and Communication-Efficient VFL. This investigation centers on how each technique's unique strategy for leveraging unaligned data and managing inter-party communication serves to overcome critical bottlenecks in conventional VFL.

Review

FedCVT: Semi-supervised Vertical Federated Learning with Cross-view Training:

* **Key Goals:** The primary goal of FedCVT is to improve the performance of VFL models, especially in realistic scenarios where the number of perfectly aligned (overlapping) samples between participating parties is small. It aims to solve the problem of valuable, non-aligned data being left unused during training.
* **Challenges Addressed:** The paper addresses the critical limitation of standard VFL, which requires a large number of aligned samples to achieve good performance. It recognizes that in many real-world collaborations (e.g., between a bank and a retail company), the sample overlap is often limited, making traditional VFL less effective.
* **Key Contributions:** It introduces a semi-supervised learning framework that effectively utilizes both the limited aligned data and the much larger set of non-aligned data held by each party.
* It proposes a novel method for leveraging this data:
  1. **Representation Estimation:** It estimates the missing feature representations for the non-aligned samples.
  2. **Pseudo-Labeling:** It generates high-confidence "pseudo-labels" for the now-complete but unlabeled samples, effectively expanding the training dataset.
  3. **Cross-View Training:** It trains three classifiers jointly (one for each party's "view" and one for the combined "view") to improve the model's overall representation learning and final performance.
* The method is designed to be privacy-preserving by only requiring the exchange of intermediate representations and gradients, not raw data or model parameters.

Communication-Efficient VFL with Limited Overlapping Samples (Sun et al.)

* **Key Goals:** This paper aims to simultaneously solve two of the most significant bottlenecks in practical VFL: the **extremely high communication cost** from iterative training and the **poor model performance** resulting from a limited number of overlapping samples.
* **Challenges Addressed:** It tackles the inefficiency of traditional VFL methods (like SplitNN) that require clients and the server to communicate in every single training iteration, which is slow and expensive. It also directly addresses the same "limited overlap" problem as FedCVT, noting that leaving either the communication or the data limitation problem unsolved hinders real-world VFL deployment.
* **Key Contributions:**
* It proposes **One-Shot VFL**, a framework that drastically reduces the communication between clients and the server to a single round (two uploads, one download).
* It introduces a novel mechanism for local training. Clients use partial gradients received from the server to create temporary pseudo-labels for their overlapping data via k-means clustering. This allows them to conduct effective local semi-supervised learning (SSL) using their vast unaligned data without further server communication.
* It proposes an extension called

**Few-Shot VFL**, which adds just one more communication round to intelligently expand the clients' labeled datasets, further boosting accuracy in scenarios with very few overlapping samples.

* The combined approach leads to a massive reduction in communication cost (over 330x reported on CIFAR-10) while simultaneously achieving significant accuracy gains (over 46.5%) compared to previous methods.

Comparison Table:

|  |  |  |
| --- | --- | --- |
| Metric | FedCVT | Communication-Efficient VFL |
| Communication Overhead | **Iterative**. Requires exchanging intermediate representations and gradients in each training round. This is more efficient than sharing raw data but still suffers from the high costs of iterative communication | **Minimal (One-Shot or Few-Shot)**. Designed to be extremely low. One-shot VFL requires only one round of communication (2 uploads, 1 download) for the entire process. Few-shot VFL adds just one more round. This is its primary advantage. |
| Label Requirement | **Single Party with Labels**. Assumes a typical VFL setup where only one party (Party A) possesses the ground-truth labels for the data | **Server with Labels**. Assumes the labels for the overlapping samples are held by a central server, and the clients themselves are unlabeled. |
| Sample Overlap Assumption | **Limited Overlap**. The entire method is built on the assumption that the set of aligned samples is small, and its main goal is to leverage the large pool of non-aligned samples. | **Limited Overlap**. This is also a core assumption. The method is designed to work with a small set of overlapping samples (Xo ) and a large set of client-specific, unaligned samples (Xu ) |
| Architecture Components | **Two Parties (A and B)**. Party A holds the labels. The system uses multiple models: local neural networks on each party to create representations (hu ,hc ) and three distinct classifiers (fA,fB,fAB) for cross-view training | **K Clients and a Server**. A classic VFL setup where clients hold feature extractors (fk ) and the server holds the main classifier (fc ) |
| Use of Unlabeled Data | **Representation Estimation & Pseudo-Labeling**. It uses the aligned data to learn how to *estimate* the missing feature parts for the non-aligned data. It then uses the complete model to generate high-confidence pseudo-labels for these samples to expand the training set. | **Clustering Gradients & Local SSL**. In one-shot VFL, it uses gradients from the server to cluster and create temporary labels for overlapping data. This labeled seed set is then used to kickstart a local semi-supervised learning process that leverages all of the client's unaligned data. |

## Part C

Discussion & Analysis