**Comparative Study and Implementation of Communication-Efficient and Semi-Supervised Vertical Federated Learning**

**Distributed Machine Learning**

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

**Abstract**This work studies two modern Vertical Federated Learning (VFL) approaches: FedCVT (semi-supervised cross-view training) and communication-efficient VFL under limited overlapping samples (one-shot/few-shot) by Sun et al. The methods aim to overcome small alignment between parties and high communication costs in VFL. After reviewing goals, assumptions, and architectures, a simplified two-client VFL simulation is implemented to illustrate design trade-offs. The discussion clarifies the implications of limited overlap, mechanisms for exploiting unlabeled data, and practical recommendations for low-resource deployments.

**Introduction**Vertical Federated Learning (VFL) enables organizations that hold different feature spaces for partially overlapping entities to collaboratively train models without sharing raw data. Two major practical hurdles are limited overlapping samples across parties and high communication overhead from iterative exchanges of intermediate representations and gradients. FedCVT addresses the data utilization problem via cross-view semi-supervised training that can exploit unlabeled and unaligned samples. In contrast, Sun et al. propose one-shot and few-shot VFL protocols that dramatically reduce communication while retaining accuracy when overlap is very small. This report compares both approaches and demonstrates a toy simulation reflecting each method’s design choices.

Background:

VFL and practical challenges

- **Limited overlap**: Entity resolution across organizations is often partial, producing small aligned sets and large unaligned pools that ordinary VFL may not exploit well.  
- **Communication cost**: Multi-round exchanges of representations and gradients can be costly or infeasible in bandwidth- or latency-constrained environments, motivating algorithms that reduce rounds.

**Part A: Paper review and comparison**

**FedCVT:**

- **Goal**: Improve VFL when aligned samples are scarce by exploiting unlabeled and unaligned data through cross-view semi-supervised training.  
- **Core** challenge: Vanilla VFL relies on aligned, labeled samples; unlabeled and unaligned data remain underused.  
- **Contributions**:  
(i) Estimation of missing-view representations,  
(ii) Pseudo-labeling with confidence/consensus selection to expand training data,  
(iii) Joint training of three classifiers (per-view and fused) to reinforce cross-view consistency while sharing only masked intermediate representations/gradients.  
  
**How FedCVT works (essentials)**

(i) Learn view-specific encoders;

(ii) Estimate missing-view representations

(iii) Generate pseudo-labels and filter by confidence/consensus

(iv) Jointly train per-view and fused classifiers on the expanded dataset in a VFL coordination loop.  
- Empirical findings: Across multiple datasets (e.g., NUS-WIDE, Vehicle, CIFAR-10), the method improves upon vanilla VFL constrained to overlaps; ablations support the utility of representation estimation and pseudo-labeling.  
  
**Communication-efficient VFL (Sun et al.):**   
- **Goal**: Reduce communication and retain accuracy under very limited overlap via one-shot VFL and a few-shot refinement.  
- **Core challenge**: Many VFL methods require numerous rounds and cannot exploit unaligned data effectively; small overlap harms both learning signal and efficiency.  
- **Contributions**:  
1) One-shot VFL: clients perform local SSL (on overlapped and unaligned data) to learn strong feature extractors and interact with the server minimally (two uploads, one download total),  
2) Few-shot VFL: adds one extra round to further boost accuracy when overlap is tiny,  
3) Extensive experiments on image and tabular benchmarks showing large communication savings versus strong baselines.  
  
How one-shot/few-shot works (essentials)  
- One-shot: Clients train local feature extractors via SSL using both overlapping and unaligned samples; the server provides minimal coordination and fusion, dramatically cutting bandwidth.  
- Few-shot: Adds a single extra round to expand the effective supervised signal and improve final accuracy when overlap is minimal.

**Comparison table (paste into Word and apply a grid table style)**

|  |  |  |
| --- | --- | --- |
| **Aspect** | **FedCVT (Kang et al.)** | **Comm.-Efficient VFL (Sun et al.)** |
| Communication | Multi-round training exchanging masked intermediate representations/gradients; heavier than one-shot | One-shot uses two uploads + one download total; few-shot adds one extra round; communication is drastically reduced. |
| Label requirement | Pseudo-labeling expands training data; effective with scarce labels | Local SSL leverages unlabeled overlapped and unaligned samples to learn feature extractors. |
| Overlap assumption | Designed for limited overlap; estimates missing-view representations to utilize non-aligned data | Explicitly targets minimal overlap; robust with one/few communication rounds. |
| Architecture components | View-specific encoders; missing-view representation estimator; pseudo-label generator; three jointly trained classifiers | Client-side SSL-trained feature extractors; minimal server coordination for one-shot/few-shot fusion. |
| Use of unlabeled data | Central via pseudo-labeling and cross-view agreement filtering. | Central via SSL on both overlapped and unaligned data at clients. |

**Part B: Conceptual simulation**

**Objective**Implement a simplified two-client, one-server VFL simulation over a 2-view dataset, with limited overlap and dummy privacy masking. Provide two modes: (a) one-shot-style (communication-efficient) and (b) FedCVT-style (pseudo-labeling/cross-view flavored).  
 **Dataset and partitioning**  
- Use any CSV/tabular dataset (e.g., UCI Adult or Credit).  
- Split columns into two disjoint feature views: A-view and B-view.  
- Define an overlap ID subset (e.g., 10–20% of rows) present at both clients.  
- Define unaligned sets for each client (rows present only at that client). **Models**- Client encoders: small MLPs producing d-dimensional embeddings per sample.  
- Server head: shallow classifier over concatenated embeddings [zA || zB] for overlapped samples.  
 **Privacy simulation**- Dummy pseudo-masking: add small Gaussian noise to embeddings before transfer.  
- No raw features leave the clients; only labels for overlapped training are used at the server.  
  
**Training modes  
  
1) One-shot-style (communication-efficient)**  
- Local SSL: each client trains encoders on its unaligned data via a self-supervised loss (e.g., autoencoding or consistency).  
- Overlap fusion: clients send masked embeddings for overlapped samples once; the server trains a shallow head on concatenated embeddings and labels.  
- Optional single download: server sends back the head (or logits) for calibration; stop after this minimal exchange.  
 **2) FedCVT-style (semi-supervised cross-view flavor)**- Initial step: as above, train the server head with overlapped masked embeddings.  
- Pseudo-labeling: use server logits to pseudo-label high-confidence extra samples; optionally require agreement with per-view predictions.  
- Missing-view approximation: use encoders to approximate the other view’s representation for unaligned samples.  
- Joint training: expand training with pseudo-labeled samples and repeat a few rounds of masked-embedding exchange to mimic cross-view training.  
 **Evaluation**- Report accuracy (or F1) on a held-out portion of overlapped samples.  
- Optionally, report communication rounds and rough embedding transfer sizes.

**Part C: Discussion and analysis**  
  
**Real-world implications of limited overlap**  
Limited overlap is common in cross-organization collaborations (finance, healthcare, retail–ads), causing the supervised signal to be concentrated in a small aligned set. Unless methods leverage unaligned data or minimize communication waste, models may underperform, and collaboration becomes impractical in bandwidth- or privacy-constrained environments.  
  
**How FedCVT exploits unlabeled data**  
FedCVT enlarges the effective training set by pseudo-labeling unlabeled samples and estimating missing-view representations so that samples lacking full features can participate in training. A confidence/consensus filter reduces noise, and joint training across per-view and fused classifiers enforces cross-view consistency.  
  
**Practicality in low-resource scenarios**  
When bandwidth and coordination are constrained, one-shot/few-shot VFL is more practical: most learning happens locally via SSL, requiring only one or a few communication rounds. When moderate communication is acceptable and large pools of unlabeled, unaligned data exist, FedCVT’s cross-view mechanisms can provide accuracy gains over vanilla multi-round VFL.  
  
**Conclusion**  
FedCVT and one-shot/few-shot VFL push VFL beyond overlap-only training. Cross-view semi-supervised learning (FedCVT) is appealing for data-rich settings with manageable communication budgets, while communication-efficient protocols are preferable for low-resource deployments with tiny overlaps.  
  
**References**  
[11] Kang et al., “FedCVT: Semi-Supervised Vertical Federated Learning with Cross-View Training,” ACM/ArXiv.  
[12] Sun et al., “Communication-Efficient Vertical Federated Learning with Limited Overlapping Samples,” ICCV 2023.