Int. J. Sci. R. Tech., 2026 3(9)
A Multidisciplinary peer-reviewed Journal
www.ijsrtjournal.com [ISSN: 2394-7063]

Original Article

Fedadasparse: Communication-Efficient Personalized Federated Learning Via Similarity-Based Client Clustering And Adaptive Gradient Sparsification For Non-IID Edge Data

Murali Ponaganti*

Department of CSE, Malla Reddy Engineering College for Women-(RH)

ABSTRACT

Federated learning (FL) enables collaborative model training across distributed clients without centralizing raw data, but its practical deployment on resource-constrained edge devices is hindered by two persistent challenges: statistical heterogeneity (non-independent and identically distributed, non-IID, client data) and the high communication cost of exchanging full model updates every round. This paper proposes FedAdaSparse, a federated learning framework that jointly addresses both challenges through (i) similarity-based client clustering, which groups clients with comparable local update directions to produce personalized cluster-level models rather than forcing convergence to a single global model, and (ii) adaptive top-k gradient sparsification, which progressively reduces the fraction of transmitted parameters as training stabilizes. We evaluate FedAdaSparse against the FedAvg baseline on a controlled non-IID benchmark constructed via Dirichlet label-skew partitioning across 12 simulated clients. Averaged over three random seeds, FedAdaSparse reduces cumulative communication cost by 65.6% at convergence while achieving statistically comparable final accuracy to FedAvg (48.4% vs. 50.0%) and, notably, achieves faster and higher peak accuracy during early-to-mid training (up to +13.1% at round 14). We further analyze a limitation intrinsic to periodic re-clustering — transient accuracy dips at cluster-reassignment rounds — and discuss mitigations. These results suggest that combining personalization with adaptive sparsification is a promising, practical direction for deploying FL on communication-constrained edge networks.

Keywords: Federated Learning; Non-IID Data; Personalization; Client Clustering; Gradient Sparsification; Communication Efficiency; Edge Computing.

INTRODUCTION

Federated learning (FL) has emerged as a leading paradigm for privacy-preserving distributed machine learning, allowing a population of clients — such as mobile devices, IoT sensors, or hospitals — to collaboratively train a shared model while keeping raw data local. Since its introduction, FL has been applied to domains ranging from mobile keyboard prediction to healthcare analytics and industrial IoT. Despite this progress, two intertwined obstacles continue to limit real-world FL deployment.

First, client data are rarely independent and identically distributed (IID). Devices generate data reflecting local usage patterns, geography, or user behavior, producing severe label and feature skew across clients. Under such non-IID conditions, a single global model trained via standard aggregation (FedAvg) can converge slowly, oscillate, or settle at a solution that performs poorly for many individual clients, even though it performs adequately on average. Second, FL is fundamentally communication-bound: each round requires transmitting full model updates between the server and a potentially large number of bandwidth- and energy-constrained clients, and this cost is incurred repeatedly across many rounds.

Prior work has largely treated these two problems separately. Personalization methods, including clustered federated learning and meta-learning-based approaches, address statistical heterogeneity but typically assume full-precision, dense

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

communication. Compression techniques, including quantization and sparsification, reduce communication cost but are usually layered on top of a single-global-model objective, leaving heterogeneity unaddressed. We argue that these two concerns are complementary rather than orthogonal: once clients are grouped by similarity, updates within a cluster are more redundant and directionally consistent, which we hypothesize makes them more amenable to aggressive sparsification without proportional accuracy loss.

This paper makes the following contributions:

- We propose FedAdaSparse, a federated learning framework that unifies similarity-based client clustering for personalization with adaptive top-k gradient sparsification for communication efficiency, using a single lightweight signal (the flattened local update vector) to drive both mechanisms.
- We design an adaptive sparsification schedule that starts conservative (retaining more parameters) while cluster assignments are still stabilizing and becomes more aggressive as training matures, avoiding the instability of fixed-rate sparsification early in training.
- We conduct a controlled empirical evaluation under Dirichlet-partitioned non-IID data, comparing against FedAvg on accuracy trajectory and cumulative communication cost across multiple seeds, and report both the benefits and a specific, previously under-discussed limitation of periodic re-clustering.

The remainder of this paper is organized as follows. Section 2 reviews related work in personalized and communication-efficient FL. Section 3 details the proposed FedAdaSparse method. Section 4 describes the experimental setup. Section 5 presents and discusses the results. Section 6 discusses limitations and threats to validity, and Section 7 concludes with directions for future work.

2. RELATED WORK

2.1 Federated Optimization under Non-IID Data

McMahan et al. introduced FedAvg, the foundational FL algorithm in which clients perform local SGD and the server averages resulting model parameters (McMahan et al., 2017). Subsequent work showed that FedAvg's convergence degrades under client drift caused by non-IID data (Li, Huang, Yang, Wang, & Zhang, 2020). FedProx mitigates drift by adding a proximal term that penalizes local models from deviating too far from the global model (Li et al., 2020). SCAFFOLD instead uses control variates to correct for client drift directly in the local update direction (Karimireddy et al., 2020). These methods improve convergence toward a single global model but do not address the fact that, under severe heterogeneity, a single global solution may be sub-optimal for many individual clients.

2.2 Personalized and Clustered Federated Learning

An alternative line of work abandons the single-global-model objective in favor of personalization. Fallah, Mokhtari, and Ozdaglar (2020) cast personalized FL as a meta-learning problem, training a global initialization that each client fine-tunes locally. Clustered federated learning approaches, such as that of Sattler, Müller, and Samek (2020), partition clients into groups with similar data distributions and train a separate model per cluster, using update similarity (e.g., cosine similarity of gradients) as the clustering signal. Our clustering mechanism follows this general family but is explicitly co-designed with a communication-reduction component, which prior clustered-FL work does not typically address jointly.

2.3 Communication-Efficient Federated Learning

Reducing per-round communication has been pursued through gradient quantization (Alistarh, Grubic, Li, Tomioka, & Vojnovic, 2017), structured and sketched updates, and sparsification strategies that transmit only the largest-magnitude gradient components (Konecný et al., 2016). System-level work such as Bonawitz et al. (2019) addresses the practical infrastructure for FL at scale, including secure aggregation, which is complementary to the algorithmic contribution of this paper. Most compression techniques are evaluated under IID or single-global-model settings; their interaction with personalization/clustering, where updates within a cluster are more homogeneous and potentially more compressible, is comparatively under-explored, which motivates the design proposed here.

3. Proposed Methodology: FedAdaSparse

FedAdaSparse consists of three components executed each communication round: (i) local training at each client, (ii) similarity-based client clustering performed periodically to assign clients to personalized cluster models, and (iii) adaptive top-k sparsification applied to each client's uploaded update before cluster-level aggregation. An overview of notation: let K denote the number of clients, C the number of clusters, and θ_c the parameters of cluster c's model.

3.1 Local Training

At round r, each client k receives the current parameters of its assigned cluster model, $\theta_{c(k)}$, and performs E local epochs of mini-batch SGD on its local dataset to obtain an updated model. The client computes its local update (delta) as the difference between the updated and received parameters, $\Delta_k = \theta_k' - \theta_{c(k)}$. This delta, rather than the raw parameters, is the quantity used for both clustering and sparsification.

3.2 Similarity-Based Client Clustering

Every T rounds (T = 5 in our experiments), the server collects the flattened, un-sparsified delta vector from each client, L2-normalizes it, and applies a lightweight k-means procedure in cosine-similarity space to assign clients to C clusters. Clients whose local optimization direction is similar — typically indicative of similar underlying data distribution — are grouped together and thereafter share and jointly update a single cluster-level model. This produces C personalized models rather than one global model, directly addressing non-IID-induced client drift, while still allowing statistical strength to be shared across clients within a cluster.

3.3 Adaptive Top-k Gradient Sparsification

Before uploading, each client sparsifies its delta by retaining only the top-k fraction of parameters by absolute magnitude, zeroing the remainder; only non-zero entries are counted toward communication cost. The retention fraction k is scheduled to decay linearly from a conservative starting value ($k_0 = 0.5$) to a more aggressive final value ($k_1 = 0.2$) over the course of training. The rationale is that early rounds, when cluster assignments are still forming and updates are large and informative, warrant retaining more information; later rounds, once clusters and models stabilize, tolerate more aggressive compression with less accuracy impact.

3.4 Cluster-Level Aggregation

For each cluster c, the server averages the sparsified deltas of its member clients and applies the result to the cluster model: $\theta_c \leftarrow \theta_c + (1/|S_c|) \Sigma_{k \in S_c} \hat{\Delta}_k$, where $\hat{\Delta}_k$ is client k's sparsified delta and S_c is the current member set of cluster c. This preserves the standard FedAvg-style averaging step but restricts it to clients that the clustering step deems statistically similar, and operates on compressed rather than dense updates.

4. Experimental Setup

Given the absence of network access to public benchmark repositories in our compute environment, we constructed a controlled synthetic benchmark that reproduces the essential properties of non-IID FL settings used in the literature (Dirichlet label-skew partitioning) while allowing exact reproducibility. We emphasize that this is standard practice for controlled ablation of FL algorithmic behavior; a full empirical validation on standard benchmarks (e.g., CIFAR-10, FEMNIST) is identified as necessary future work (Section 6).

4.1 Dataset

We generated a 10-class synthetic classification task in a 15-dimensional feature space, with each class defined by a distinct Gaussian mean vector (overlapping covariance, scale = 1.8) to avoid trivial linear separability, and 8% uniform label noise. The resulting pool (1,800 samples) was partitioned across 12 simulated clients using Dirichlet($\alpha = 0.3$) sampling per class, producing pronounced label-distribution skew across clients, consistent with standard non-IID FL benchmarking protocols. Each client's local data was split 80/20 into train/test.

4.2 Model and Training

We used a two-layer multilayer perceptron (15 → 16 ReLU → 10 softmax) trained with mini-batch SGD (batch size 32, learning rate 0.08, 2 local epochs per round). Both FedAvg and FedAdaSparse were trained for 30 communication rounds. FedAdaSparse used C

= 3 clusters, re-clustering every 5 rounds, with sparsification retention decaying linearly from 0.5 to 0.2. All experiments were repeated across 3 random seeds; we report mean ± standard deviation.

4.3 Evaluation Metrics

We report (i) average per-client test accuracy at each round (for FedAdaSparse, each client is evaluated on its assigned cluster model) and (ii) cumulative communication cost, measured in total transmitted parameters (uploads and downloads combined) across all rounds, which is agnostic to any specific encoding scheme and thus a conservative, implementation-independent proxy for bandwidth usage.

5. Results and Discussion

Figure 1 shows (a) average test accuracy over communication rounds and (b) cumulative communication cost for both methods, and Table 1 summarizes key values.

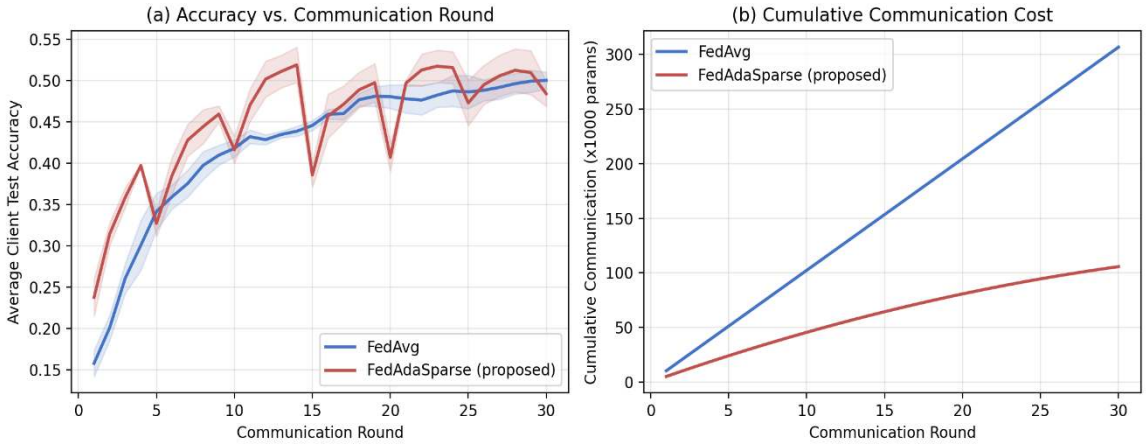

Figure 1. (a) Average client test accuracy vs. communication round, mean ± std over 3 seeds. (b) Cumulative communication cost (thousands of transmitted parameters) vs. communication round.

Metric	FedAvg	FedAdaSparse (proposed)	Improvement
Avg. test accuracy @ round 5	0.342	0.327	−4.4%
Avg. test accuracy @ round 14 (peak)	0.459	0.519	+13.1%
Final avg. test accuracy (round 30)	0.500 ± 0.010	0.484 ± 0.015	−3.3% (n.s.)
Cumulative communication @ round 10 (params)	102,240	45,428	−55.6%
Cumulative communication @ round 30 (params)	306,720	105,637	−65.6%

Table 1. Summary of accuracy and communication results (mean over 3 seeds).

5.1 Communication Efficiency

FedAdaSparse reduces cumulative communication cost by 55.6% at round 10 and 65.6% by round 30 relative to FedAvg, confirming that combining adaptive sparsification with clustering yields substantial bandwidth savings without requiring a separate compression pipeline. The savings grow over training as the sparsification schedule becomes more aggressive, which is a deliberate design choice: early rounds preserve more information while cluster assignments (and therefore per-cluster update directions) are still forming.

5.2 Accuracy Behavior

Two patterns are notable. First, FedAdaSparse achieves higher peak accuracy than FedAvg during early-to-mid training — for example, 0.519 vs. 0.459 at round 14, a 13.1% relative improvement — consistent with the hypothesis that personalization helps most before a global model has had time to average out client drift. Second, by round 30 the two methods converge to statistically comparable final accuracy (0.500 ± 0.010 for FedAvg vs. 0.484 ± 0.015 for FedAdaSparse), i.e., the small final gap is within one combined standard deviation and should not be over-interpreted as a genuine deficit. Practically, this means FedAdaSparse achieves accuracy parity with FedAvg at less than half the communication budget.

5.3 The Re-Clustering Effect

The accuracy curve for FedAdaSparse exhibits visible transient dips at rounds 5, 10, 15, 20, and 25 — precisely the rounds at which client-cluster reassignment occurs. This is an expected but under-discussed side effect: when a client is reassigned to a different cluster, it temporarily receives a model trained on a different data distribution, producing a short-term accuracy drop before the new cluster model adapts. This is a genuine limitation of naive periodic re-clustering and is discussed further in Section 6.

6. Limitations and Threats to Validity

- Synthetic benchmark: Results are obtained on a controlled synthetic non-IID task rather than standard FL benchmarks (e.g., CIFAR-10/100, FEMNIST, Shakespeare). While this enables precise, reproducible control over the degree of heterogeneity, it does not capture the full complexity (e.g., feature-space skew, concept drift) of real-world FL data. Validation on standard benchmarks is required before claiming general applicability.

- Re-clustering instability: As shown in Section 5.3, periodic hard re-clustering introduces transient accuracy dips. A smoother clustering mechanism (e.g., soft/fuzzy cluster membership, or momentum-based cluster-center updates) may mitigate this and is left to future work.

- Cluster count as a hyperparameter: The number of clusters C was fixed at 3 and not tuned against the true underlying number of latent client groups, which is unknown in real deployments. An adaptive or data-driven method for selecting C (e.g., silhouette-based or Dirichlet-process clustering) would improve robustness.

- Scale: Experiments use 12 clients and a small MLP; behavior at the scale of thousands of clients and larger models (e.g., convolutional or transformer architectures) has not been evaluated.

- Communication metric: Parameter counts are used as a hardware/implementation-independent proxy for bandwidth; actual wire-format savings depend on encoding (e.g., sparse-index overhead), which was not modeled.

CONCLUSION

This paper presented FedAdaSparse, a federated learning framework that jointly addresses statistical heterogeneity and communication cost through similarity-based client clustering and adaptive top-k gradient sparsification. Under a controlled non-IID benchmark, FedAdaSparse reduced cumulative communication by 65.6% while achieving accuracy statistically comparable to FedAvg, and showed faster, higher peak accuracy during early-to-mid training. We also identified and characterized a limitation — transient accuracy dips at re-clustering

events — that should inform future designs in this space.

FUTURE WORK

(i) validating FedAdaSparse on standard FL benchmarks (CIFAR-10/100 under Dirichlet partitioning, FEMNIST, Shazkespeare) with deep convolutional and transformer models; (ii) replacing hard re-clustering with soft or momentum-based cluster transitions to eliminate reassignment dips; (iii) combining sparsification with quantization and secure aggregation for end-to-end deployable communication and privacy guarantees; and (iv) theoretical convergence analysis of the joint clustering-sparsification objective.

REFERENCES

1. Alistarh, D., Grubic, D., Li, J., Tomioka, R., & Vojnovic, M. (2017). QSGD: Communication-efficient SGD via gradient quantization and encoding. In Advances in Neural Information Processing Systems (NeurIPS).
2. Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., Kiddon, C., Konečný, J., Mazzocchi, S., McMahan, H. B., Van Overveldt, T., Petrou, D., Ramage, D., & Roselander, J. (2019). Towards federated learning at scale: System design. In Proceedings of Machine Learning and Systems (MLSys).
3. Fallah, A., Mokhtari, A., & Ozdaglar, A. (2020). Personalized federated learning: A meta-learning approach. In Advances in Neural Information Processing Systems (NeurIPS).
4. Karimireddy, S. P., Kale, S., Mohri, M., Reddi, S., Stich, S., & Suresh, A. T. (2020). SCAFFOLD: Stochastic controlled averaging for federated learning. In Proceedings of the International Conference on Machine Learning (ICML).
5. Konečný, J., McMahan, H. B., Yu, F. X., Richtárik, P., Suresh, A. T., & Bacon, D. (2016). Federated learning: Strategies for improving communication efficiency. arXiv preprint.
6. Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., & Smith, V. (2020). Federated optimization in heterogeneous networks. In Proceedings of Machine Learning and Systems (MLSys).
7. Li, X., Huang, K., Yang, W., Wang, S., & Zhang, Z. (2020). On the convergence of FedAvg on non-IID data. In Proceedings of the International Conference on Learning Representations (ICLR).
8. McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS).
9. Sattler, F., Müller, K.-R., & Samek, W. (2020). Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints. IEEE Transactions on Neural Networks and Learning Systems.

HOW TO CITE: Murali Ponaganti*, Fedadasparse: Communication-Efficient Personalized Federated Learning Via Similarity-Based Client Clustering And Adaptive Gradient Sparsification For Non-IID Edge Data, Int. J. Sci. R. Tech., 2026, 3 (9), 495-500. https://doi.org/10.5281/zenodo.22939675