FIELDING: Clustered Federated Learning with Data Drift

Minghao Li[1] · Dmitrii Avdiukhin[2] · Rana Shahout[1] · Nikita Ivkin[3] · Vladimir Braverman[4,5] · Minlan Yu[1]

1: Harvard University, 2: Logical Intelligence, 3: Amazon, 4: Johns Hopkins University, 5: Google

MOTIVATION

- Federated Learning (FL) enables collaborative training without sharing raw data.
- **Client heterogeneity** slows convergence and reduces accuracy.
- Clustered FL (CFL) improves performance—but suffers under **data drift**.

PROBLEM: DATA DRIFT

Client local data evolves over time:

- **Label shift:** changing label frequencies.
- **Covariate shift:** changing input distributions.
- **Concept drift:** changing input–output relationships.
  - ◦ Synchronized drift: all clients share a new $P(y|x)$.
  - ◦ Distributed drift: clients drifts to different $P(y|x)$.

Consequence: clusters become outdated → worse performance.

(a)

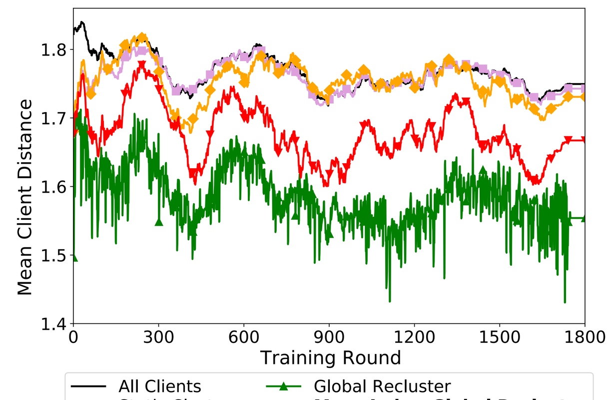

Figure 1(a): Global re-clustering decreases label distribution L1 distance

(b)

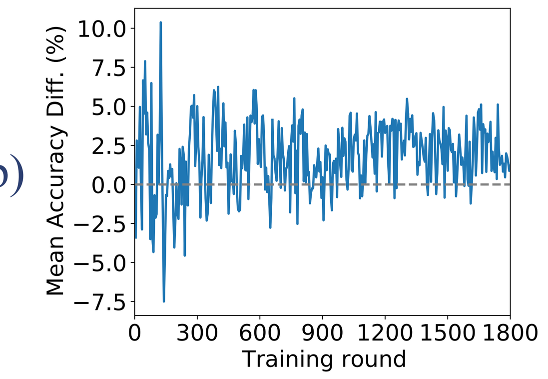

Figure 1(b): Combined approach vs. global re-clustering. Positive numbers mean the combined approach is better.

LIMITATIONS of EXISTING DRIFT HANDLING

Per-client adjustment:

- ✅ Efficient for small-scale drift.
- ❌ Fails under large-scale drift.

Global re-clustering:

- ✅ Handle all drifts.
- ❌ High costs + unstable accuracy.

Takeaway: Need to **balance accuracy + efficiency**.

KEY IDEAS

Combine **per-client adjustment** with **selective global re-clustering**

- ✅ Adapt efficiently to small-scale drift.
- ✅ Remain robust under large-scale drift.
- ✅ Avoid unnecessary global re-clustering + accuracy dips.

FIELDING FRAMEWORK

1. Drift Detection
- Clients monitor local data changes.

2. Per-client Adjustment
- Drifted clients move to nearest cluster.
- Low overhead, fast update, stable accuracy.

3. Selective Global Re-Clustering

Triggered only when needed:
- Measure intra-cluster heterogeneity.
- If threshold **Δ exceeded, re-cluster all clients**.

4. Agnostic to Client Representations

Enables addressing all types of data drift:
- Label distribution and embedding vectors: low overhead; handle all but distributed concept drift.
- Gradient: expensive; handle all drifts.

5. Adaptive Threshold (Δ)
- Dynamically adjusts to drift patterns.
- Multiplicative increase under consecutive global re-clustering (fast reaction to overly low Δ).
- Linear decrease to maintain the correct Δ and eventually adapts if Δ needs to be decreased.

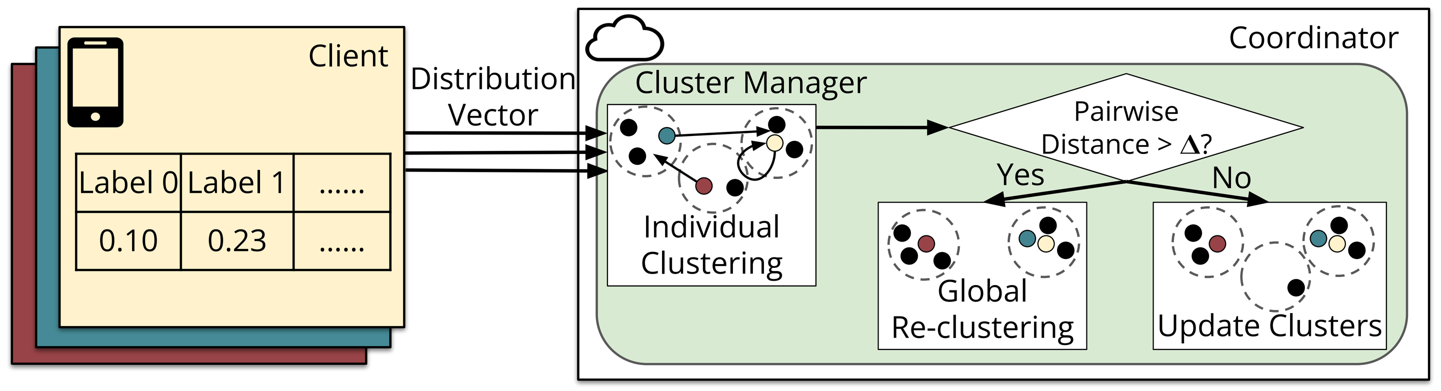

Figure 2: Clustering in FIELDING with Label Distribution as Representations.

TRAINING WORKFLOW

1. Detect drifted clients.
2. Reassign to nearest cluster.
3. Check intra-cluster heterogeneity (compare with Δ).
4. Trigger global re-clustering if needed.
5. Train cluster-specific models.

RESULTS

Setup:
- Datasets: Functional Map of the World (FMoW), Cityscapes, Waymo Open, Open Images.
- Models: ResNet-18, ShuffleNet v2, VisionTransformer-B16.

Performance Gains:
- **2.4% – 6.9%** accuracy improvement.
- **1.38× – 3.10×** faster convergence.
- Up to **24% better accuracy than static clustering**.

Time-to-Accuracy:
- Faster convergence across datasets.
- More stable under data drift.

Takeaways:
- Handles **all types of data drift**.
- Re-clusters **all drifted clients** (not just selected ones).
- Uses lightweight client representations.
- Avoids unnecessary global re-clustering.

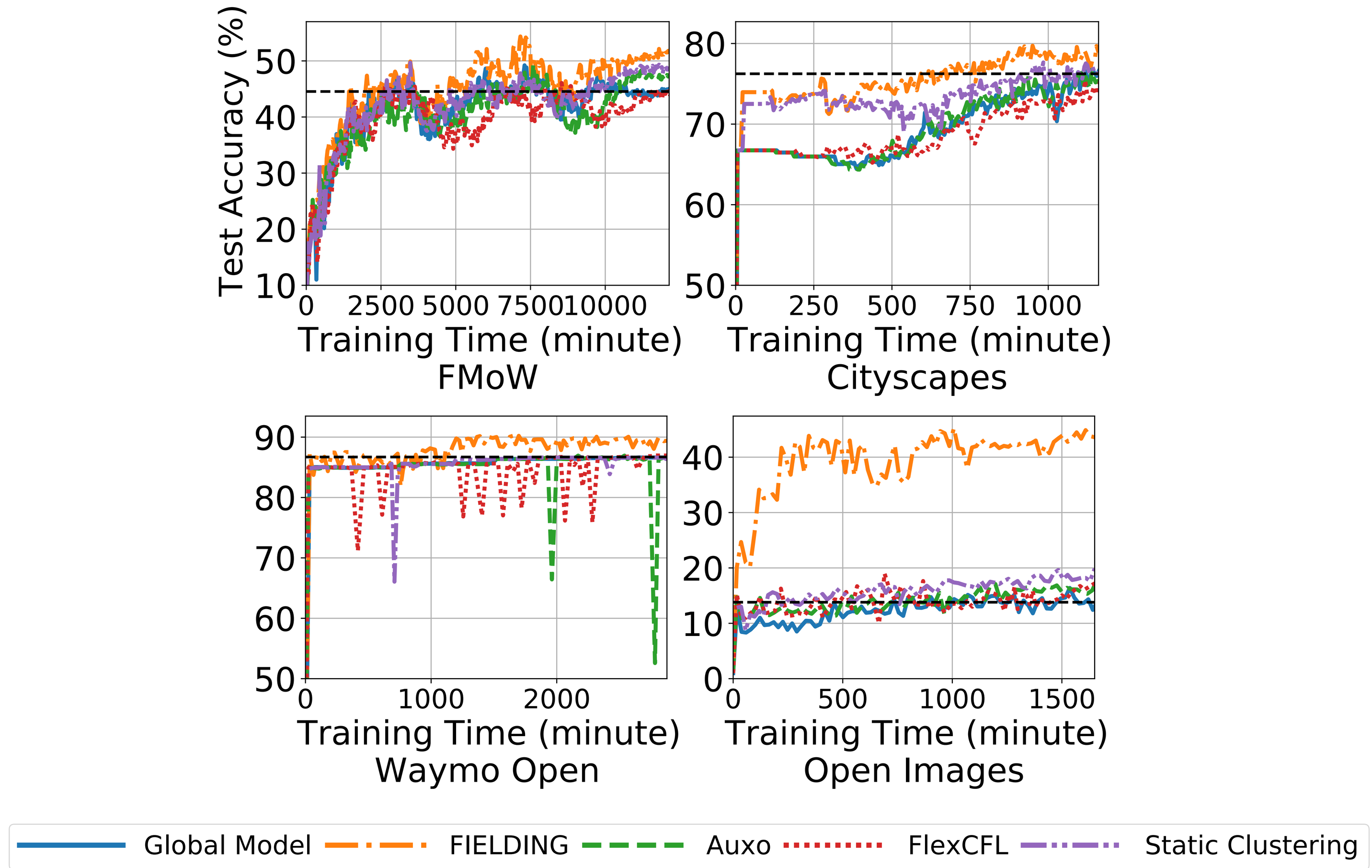

Figure 3: Time to Accuracy (TTA) Comparison (FIELDING vs. Non-Clustering vs. Auxo vs. FlexCFL vs. Static Clustering).

CONCLUSION

- Data drift is a key challenge in real-world FL.
- FIELDING provides a robust and efficient solution under dynamic data drift.
- Achieves **higher accuracy, faster convergence, and performance stability**.

Harvard John A. Paulson School Of Engineering and Applied Sciences · LOGICAL INTELLIGENCE · amazon · JOHNS HOPKINS UNIVERSITY · Google