Federated Learning under Distributed Concept Drift

Ellango Jothimurugesan*, Kevin Hsieh^, Jianyu Wang*, Gauri Joshi*, Phillip B. Gibbons*

*Carnegie Mellon University  ^Microsoft
Traditional perspective: Train then deploy

In Federated Learning,
- Clients continually compute updates to the model with new data (test-then-train)
- Server averages local updates
Federated Learning: Drifts Are Staggered

- Data drawn from distribution $P_{c}^{(t)}(x, y)$ at client $c$ and time $t$
- Concept drift: $P_{c}^{(t)}(x, y) \neq P_{c}^{(t-1)}(x, y)$

Challenges:
- Drift occurs at different clients at different times
- Multiple concepts arise simultaneously

Objective: High accuracy on test data at each client, at each time step (round)
Centralized Adaptation has Low Accuracy

- **Ex:** Single concept change on SEA dataset staggered over time

  - **Locally,** drift occurs abruptly
  - **Globally,** performance degrades slowly & drift is harder to detect

- During transition, training over a mixture of concepts converges slowly

![Graph showing test accuracy over time for different methods: FedAvg (No adaptation), DriftSurf (Detection), AUE (Ensemble).]
Train Multiple Models for High Accuracy

- Ex: Single concept change on SEA dataset staggered over time
  - “Oracle” access to matrix, trains a model specialized for each concept
  - Our solution, FedDrift, trains multiple models over a learned (time-varying) clustering of clients
FedDrift Learns the Clustering of Clients

- Ideally, clusters correspond 1-to-1 with concepts
  - Avoid a single cluster for multiple concepts (“model poisoning”)
  - Avoid multiple clusters for a single concept (not collaborating when possible)
- FedDrift adaptively determines the appropriate number of clusters

1) **Cluster splitting:**
   Isolate clients via local drift detection

2) **Cluster merging:**
   Hierarchically merge clusters at the same concept (distances determined by running drift detection across clients)
Evaluation: 4 Concept Pattern

- Average accuracy of FedDrift (94.1%) close to Oracle (94.6%)
Evaluation: Functional Map of the World

- Real-world distributed drift in Functional Map of the World from WILDS
- Globally the drift is small compared to local drift for Africa
- FedDrift outperforms the best baseline (64% to 58%)
Conclusion

• Our work is the first to study data heterogeneity both over time and across clients in federated learning

• Existing centralized solutions fail on staggered drifts

• FedDrift achieves high accuracy on variety of drifts
  • Comparable to an idealized oracle algorithm on synthetic datasets
  • Outperforms the best baseline (64% to 58%) on the real-world FMoW