Event-Based Federated Q-Learning

Guener Dilsad ER Michael Muehlebach
Max Planck Institute for Intelligent Systems

MAX-PLANCK-GESELLSCHAFT

MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS

Summary

- **Objective:** Learning a policy that performs well across all environments while maintaining data privacy among the agents.
- **Challenge:** Communication overhead in Federated RL.
- **Solution:** Event-Based Federated Q-Learning (EBQAvg) algorithm.
  Key mechanism: Agents communicate updates only if there are significant changes.
- **Theoretical Analysis:** Trade-off between communication efficiency, environment heterogeneity (different state-action pairs and state transitions) and convergence rate.
- **Empirical Results:** Tested in Windy Cliff and Cart Pole environments.
  Our approach significantly reduces communication (typically around 70-80%) without sacrificing performance.
- **Impact:** Enables efficient, scalable federated reinforcement learning.

Algorithm 1 Event-Based QAvg Algorithm (EBQAvg)

Require: Number of agents n, number of rounds T, learning rate λ_t, discount factor γ, communication threshold δ, number of local updates E

Initialize Q-tables Q^k for each agent k

for $t = 1$ to T **do**

  for $k = 1$ to n **do**

    Receive broadcasted $Q_t^k \leftarrow \bar{Q}_t$

    Perform E local updates of Q_t^k with standard **Q-Learning update**

$$Q_{t'+1}^k(s,a) \leftarrow (1-\lambda_t)\, Q_{t'}^k(s,a) + \lambda_t[R(s,a) + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_k(s'|s,a) \max_{a' \in \mathcal{A}} Q_{t'}^k(s',a')]$$

    Send Q_{t+1}^k if **communication event** is triggered $|Q_{t+1}^i - Q_{[t]}^i| > \delta$, where $Q_{[t]}^i$ denotes the value of Q^k that has been last communicated

  end for

  Global Aggregation of Q-tables from all agents at server $\bar{Q}_{t+1} = \frac{1}{n}\sum_{i=1}^n Q_{[t+1]}^i$

  Broadcast $\bar{Q}_t$ to all agents

end for

Event-Based Communication in Federated Learning

Event-based communication reduces communication by triggering updates only when necessary and is robust to heterogeneous data-distributions among agents and communication failures.

Event-based communication is also effective in a distributed learning setting, where the aim is to minimize $\sum_{i=1}^n f^i(x)$, see [1]. The event-based framework reduces communication by only transmitting updates when significant changes occur. We established the following theoretical result in a convex setting,

$$|\xi_k - \xi_*|^2 \leq \kappa_P |\xi_0 - \xi_*|^2 \left(1 - \frac{\alpha}{4\kappa^{\epsilon+\frac{1}{2}}}\right)^{2k} + \frac{60\kappa^{2+2\epsilon}}{\alpha(1-|\alpha-1|)}\Delta^2,$$

where ξ_k is the server variable and ξ_* is the optimizer of $\sum_{i=1}^n f^i(x)$. Furthermore, Δ represents the error arising from the event-based communication, κ_P is a function of relaxation parameter α and condition number κ of the objective function $f(x) = \sum_{i=1}^n f^i(x)$, and ϵ is a parameter that scales the step-size of the algorithm.

We present empirical results that illustrate the effectiveness of our approach.

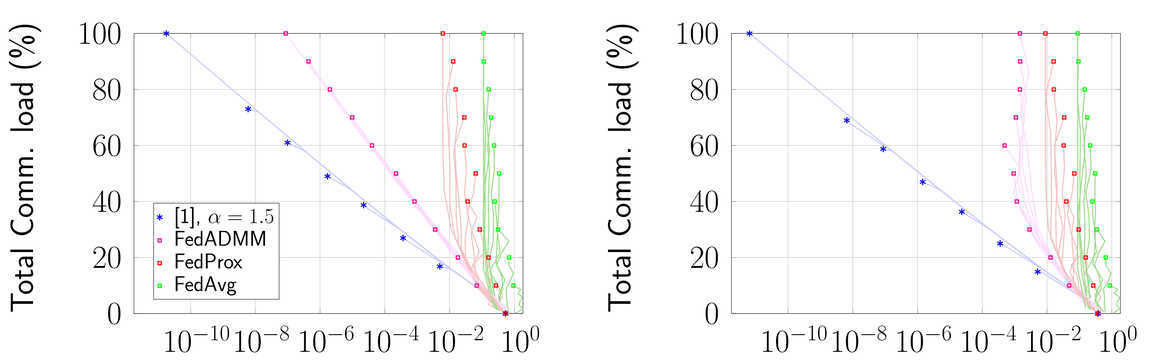

Fig. 1: The figure shows the communication load versus accuracy trade-off for different methods applied to two distinct problems: linear regression (left panel), and LASSO (right panel).

[1] G. Dilsad Er, Sebastian Trimpe, and Michael Muehlebach."Distributed Event-Based Learning via ADMM". In: arxiv:2405.10618 (2024)

Main Theorem

- **Goal of FedRL:** Enable n agents to jointly learn a policy function or a value function that performs uniformly well across all environments.
- Privacy constraints prevent agents from sharing their previous trajectories.

Optimization Problem

$$\max_{\pi} \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left\{\sum_{t=1}^{\infty} \gamma^t R(s_t,a_t) \mid s_0 \sim \mathcal{D}, a_t \sim \pi(\cdot \mid s_t), s_{t+1} \sim \mathcal{P}_i(\cdot \mid s_t, a_t)\right\}$$

where $\mathcal{D}$ is the common initial state distribution, $\mathcal{P}_i$ is the state transitions of agent i.

Theorem: Let the step-size of Algorithm 1 be $\lambda_t = \alpha$ and the discount factor be γ. Let Q_* be the Q-function of the optimal policy π^* (see above maximization).

- If the number of local updates E is chosen as $E \geq \frac{\log 2}{\alpha(1-\gamma)}$, the following inequality holds:

$$\left|\bar{Q}_t - Q_*\right|_\infty \leq \left(\frac{1}{2}\right)^t \left|\bar{Q}_0 - Q_*\right|_\infty + 2\delta + 3\epsilon.$$

  In the previous inequality, $\bar{Q}_t$ denotes the average of the distributed Q functions, $\bar{Q}_t = \frac{1}{n}\sum_{k=1}^n Q_{[t]}^k$, where $Q_{[t]}^k$, last communicated by agents $\{1,\ldots,n\}$ at iteration t. δ represents the communication threshold, and ϵ bounds the difference between Q_* and the locally optimal Q-functions, i.e., $|Q_*^k - Q_*| \leq \epsilon$.

Empirical Evaluation

Event-based communication results in a better trade-off compared to random selection of communicating agents.

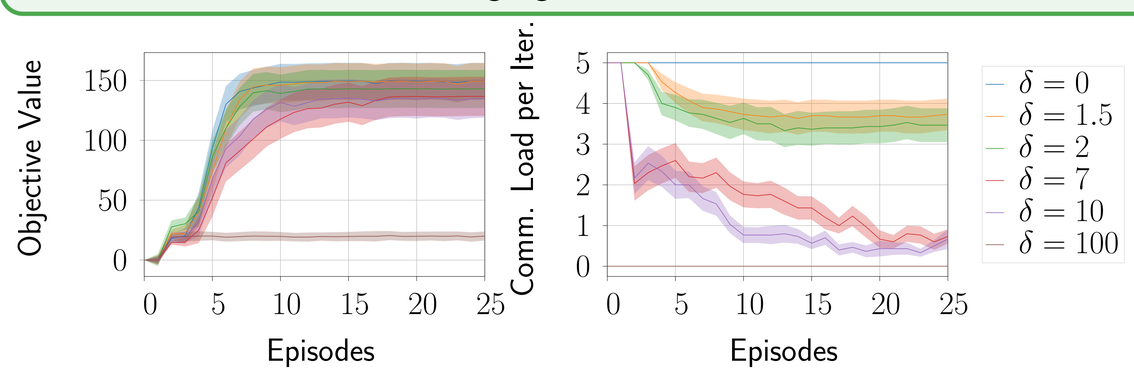

Fig. 2: Evolution of the objective over episodes for different comm. thresholds and corresponding comm. loads.

- Event-based framework significantly reduces the number of triggered communications.
- On the right, the circles represent results of event-based communication, squares represent random selection. Event-based communication achieves **better tradeoff** between communication load and objective value than random selection.

Our results highlight the potential of event-based communication in federated reinforcement learning.

By strategically transmitting updates, our scheme effectively reduces communication overhead while maintaining performance.

arxiv:[1]

OpenReview