ICML | 2022

FedNew: A Communication-Efficient and Privacy-Preserving Newton-Type Method for Federated Learning

Anis Elgabli, Chaouki Ben Issaid, Amrit Bedi, Ketan Rajawat, Mehdi Bennis, and Vaneet Aggarwal

UNIVERSITY OF OULU

6G FLAGSHIP UNIVERSITY OF OULU

CWC cwc.oulu.fi

CENTRE FOR WIRELESS COMMUNICATIONS
University of Oulu

The FL Problem : $\min_{x \in \mathbb{R}^d} \left\{ f(x) := \frac{1}{n} \sum_{i=1}^{n} f_i(x) \right\}$

Algorithms

First order approach

$$x^{k+1} = x^k - \alpha \nabla f(x^k) = x^k - \frac{\alpha}{n} \sum_{i=1}^{n} \nabla f_i(x^k)$$

+Pos

- Lower computation cost per iteration
- Lower communication cost per iteration

- Neg

- Slow convergence speed (depends on the condition number)

Second order approach: Standard Newton Method

$$x^{k+1} = x^k - \left(\sum_{i=1}^{n} \nabla^2 f_i(x^k) \right)^{-1} \sum_{i=1}^{n} \nabla f_i(x^k)$$

+Pos

- Fast converegnce
- Local quadratic convergence rate (independent on the condition number)

-Neg

- Higher computation cost per iteration
- Higher communication cost per iteration

Newton direction is the solution to:

$$z(x^k) = \arg\min_{y \in \mathbb{R}^d} \frac{1}{2} y^T \nabla^2 f(x^k) y - y^T \nabla f(x^k).$$

Our Approach :

$$\min_{y_i, y} \quad \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{2} y_i^T (\nabla^2 f_i(x^k) + \alpha I) y_i - y_i^T \nabla f_i(x^k) \right)$$

$$\text{s.t.} \quad y_i = y, \quad \forall i \in [n],$$

- One-pass ADMM to approximate the newton direction followed by inexact Newton update.
- Only one (model sized) vector is transmitted at each iteration (Low comm cost).
- Privacy is preserved. No explicit transmission of the Hessian/gradient.

1. At each client i:

$$y_i^k = (H_i^k + \alpha I + \rho I)^{-1} (g_i^k - \lambda_i^{k-1} + \rho y^{k-1})$$

2. At the PS:

$$y^k = \frac{1}{n} \sum_{i=1}^{n} (y_i^k + \lambda_i^{k-1}/\rho)$$

3. The dual variables' update

$$\boldsymbol{\lambda}_i^k = \boldsymbol{\lambda}_i^{k-1} + \rho (y_i^k - y^k).$$

4. Newton update

$$x^{k+1} = x^k - y^k,$$

Simulation Results

Fig.1

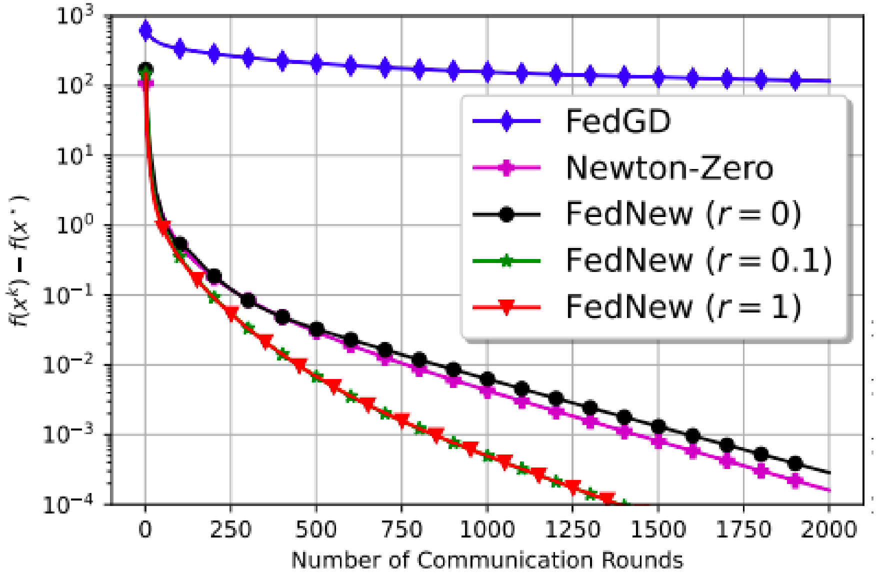
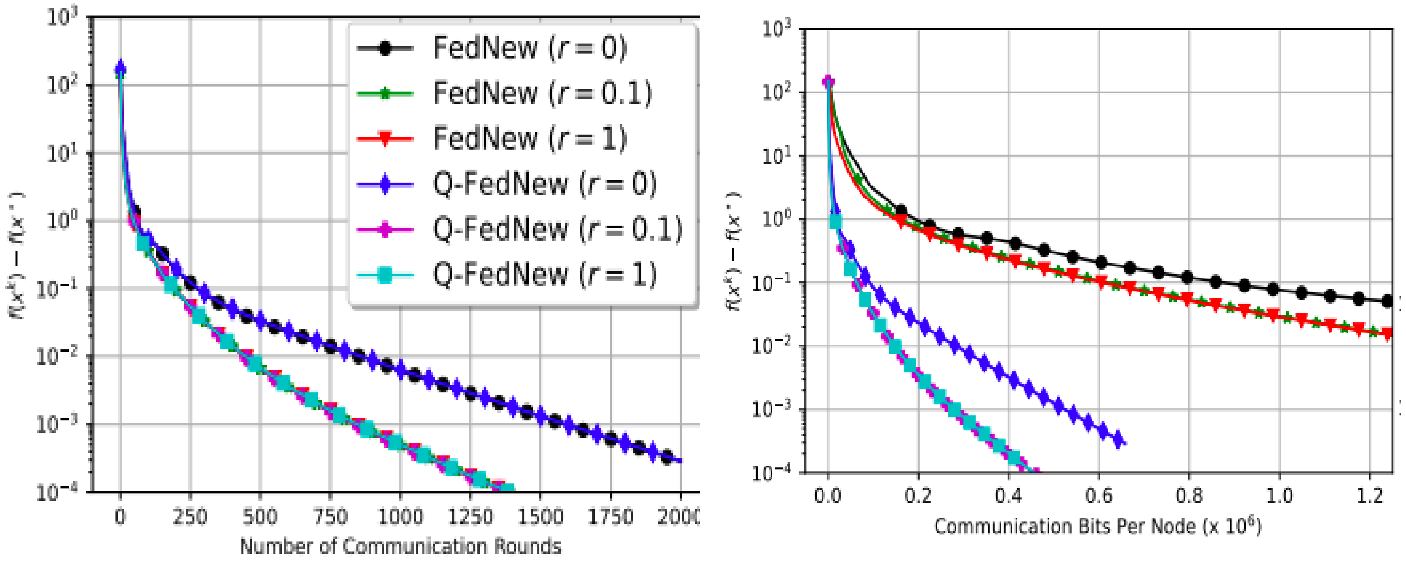


Fig.2



Dataset: a1a

Key messages from Fig 1:

- FedNew when using Zeroth hessian at each iteration (no hessian update to avoid matrix inversion, r=0) is as fast as Newton zero with better privacy gaurantees.
- FedNew when updating the hesian at each iteration (r=1) converges faster at higher computation cost.
- FedNew with hessian update at every 10-th iterations is as fast as FedNew (r=1) with 10 times less computation cost.

Key message from Fig. 2

- Quantization leads to significant reduction in terms of the number of transmitted bits while achieving the same convegence speed of FedNew.

Conclusion and Future work

- We proposed a novel communication-efficient, and privacy-preserving federated learning framework based on Newton and ADMM methods.
- The proposed approach (FedNew) ensures privacy by hiding the gradient and the Hessian information.
- FedNew achieves the same communication-efficiency of first-order methods **per iteration** while enjoying faster convergence and preserving privacy.
- FedNew is proved to follow the inexact Newton direction asymptotically.
- A further reduction in communication overhead is achieved by utilizing stochastic quantization.
- Numerical results show the superiority of FedNew compared to existing methods in terms of communication costs while ensuring privacy.
- **Future works**
  - Convergence rate analysis of FedNew and Q-FedNew.
  - Extension of the current framework to fully decentralized topology.

Questions

- For any questions, please email me at: anis.elgabli@oulu.fi