Algorithm 1: Local Gradient Aggregation

1. Initialization: \( D_j, v_j^0, v_j^0 (j = 1, 2, \ldots, N), \alpha, K, a \rightarrow \) solver
   for \( k = 0 : K \) do
     3. for \( j = 1 : N \) do
       4. Randomly shuffle the corresponding data subset \( D_j \)
       5. Compute \( \tilde{g}_j^k \)
       6. Compute \( G_j^k \)
       7. for each agent \( l, t, (j, l) \in C \) do
         8. Compute \( g_{jl}^t \)
         9. \( G_j^k = G_j^k + \tilde{g}_j^k \)
       10. \( \forall \quad G_j^k \)
       11. for \( j = 1 : N \) do
         12. Compute \( v_j^k = v_j^{k-1} + \alpha g_j^k \)
       13. end for
       14. end for
   16. end for

Figure 1 shows that LGA can maintain the high accuracy when learning from both IID and non-IID data distributions.

Figure 2 shows that LGA achieves the highest accuracy compared to the state-of-the-art methods in less number of epochs smoothly and maintains it in both IID and non-IID scenarios.

Figure 3: Average training (solid line) and validation (dash line) accuracy for different methods on (a) IID (b) non-IID data distributions for fully connected graph topology with 5 agents.

In this paper, we propose a Local Gradient Aggregation (LGA) algorithm to effectively learn from non-IID data distributions in a decentralized manner that resolves the scalability and connectivity concerns associated with using a central parameter server.

We present the convergence characteristics of the algorithm and investigate the effect of different topologies, with different combinations of agent numbers empirically.

Also, we compare the performance of LGA algorithm with state-of-the-art decentralized learning algorithms as baseline methods.

Future research directions include: (i) Computation analysis of LGA (ii) investigating projection methods other than QP (iii) empirical comparison between different extents of non-IIDness in the data distribution.

In this paper, we study two aspects of distributed deep learning: Decentralized Learning from Non-IID data, OPT 2020 workshop

• Centralized learning algorithms (e.g., federated learning) have demonstrated state-of-the-art performance in learning collaboratively from numerous agents.

• In certain use cases such as learning over a robotic network, continuous communication with a central parameter server is often not feasible. To address this concern, several decentralized learning algorithms have been proposed.

• In this paper, we study two aspects of distributed deep learning: Decentralized learning and learning from non-IID data distributions.

• We propose Local Gradient Aggregation (LGA) algorithm and show its effectiveness in learning models in a decentralized manner from both IID and non-IID data distributions.

Problem Formulation

The standard (unconstrained) empirical risk minimization problem that we are solving in decentralized distributed learning can be represented as:

\[
\min f(x) = \sum_{i=1}^{N} \sum_{x_i} f_i(x_i) = \frac{N}{n} \sum_{j=1}^{N} f_j(x_j); \text{s.t. } x_j = x_i \forall (j, l) \in C
\]

Comparing the data distribution shifts in the continual learning with the non-IID data distributions in the decentralized learning, we can leverage the techniques into the decentralized learning framework by finding the local optimal gradient for each local model.

\[
(g, \tilde{g}) := \left( \frac{\partial f(s, x)}{\partial x}, \frac{\partial f(M; x)}{\partial x} \right) \geq 0, \forall t < k.
\]


Acknowledgements: This work was partly supported by the National Science Foundation under grant number CAREER-1845969.