Summary

The technique of model averaging (MA) has not been considered for the important matrix factorization (MF) model under the scenario of federated learning (FL).

- Propose a new MA based algorithm, named FedMAvg, by judiciously combining the alternating minimization technique and MA.
- Local GD with diminishing steps and partial client participation can greatly reduce the communication cost, even under non-i.i.d. data.

Federated Matrix Factorization Model

The data samples are partitioned as \( X = [X_1, X_2, \ldots, X_P] \) and respectively owned by \( P \) distributed clients. Each client \( p \) owns non-overlapping data \( X_p \in \mathbb{R}^{m \times N_p} \), where \( N_p \) is the number of samples of client \( p \) and \( \sum_{p=1}^{P} N_p = N \).

\[
\min_{W, H} F(W, H) \triangleq \sum_{p=1}^{P} \omega_p F_p(W, H_p) \quad \text{s.t.} \quad W \in \mathbb{W}, \quad H_p \in \mathbb{H}_p, \quad \forall p \in P,
\]

where \( F_p(W, H_p) = \frac{1}{m} \Phi(X_p, WH_p) \), \( p \in P \).

Proposed FedMAvg Scheme

\( P \) could be large, \( N_p \), \( p = 1, \ldots, P \), could be unbalanced, and \( X_p, p \in P \) could be non-i.i.d.

Problem (1) is challenging to solve since it is non-convex and non-smooth, and involves two blocks of variables \( W \) and \( H \).

Algorithm Development

- Alternating Minimization:
  Given \( W^{t-1} \), each client \( p \) performs
  \[
  H_p^t = \arg \min_{H_p} F_p(W^{t-1}, H_p), \quad (2a) \\
  W_p^t = \arg \min_{W_p} F(W_p, H_p^t). \quad (2b)
  \]
  The server does \( W^t = \frac{1}{P} \sum_{p=1}^{P} \omega_p W_p^t \).

- Local GD with Diminishing Steps \( Q_2 \): - (2a) via \( Q_1 \geq 1 \) consecutive steps of PGD with respect to \( H_p \). - (2b) via \( Q_2 \geq 1 \) consecutive steps of GD with respect to \( W_p \).

Proposed FedMAvg Algorithm

\[
W^t = \frac{1}{P} \sum_{p=1}^{P} W_{p}^{t-1}
\]

and select a set of clients \( A^t \) (with size \( |A^t| = m \)) by sampling with replacement according to probabilities \( \{\omega_1, \omega_2, \ldots, \omega_P\} \), and broadcast \( W^t \) to all clients.

Client side: for client \( p \) from \( A^t \) do

- Set \( H_p^0 = H_p^{t-1} \) and \( W_p^0 = W^t \).
- for each epoch \( t = 1 \) to \( Q_2 \) do

\[
H_p^t = H_p^{t-1} - \eta \nabla_{H_p} F_p(W^{t-1}, H_p^{t-1}), \\
W_p^t = W_p^{t-1}.
\]

end for

- for each epoch \( t = 1 \) to \( Q_1 \) do

\[
W_p^t = W_p^{t-1} - \eta \nabla_{W_p} F_p(W^{t-1}, H_p^{t-1}), \\
H_p^t = H_p^{t-1}.
\]

end for

- if client \( p \) \( \notin A^t \) then

- Upload \( W_p^t \) to the server.

end if

Convergence Analysis

- Bounds:
  \[
  \|\nabla_v F_p(W, H_p) - \nabla_v F(W, H_p)\| \leq \varepsilon^2, \\
  \|\nabla_v F(W, H_p)\| \leq \varepsilon^3.
  \]
- Virtual Sequences:
  \[
  \frac{1}{Q_2} \sum_{q=1}^{Q_2} \sum_{i=1}^{Q_1} \varepsilon_i q_i W_{p}^{t+1} - \frac{1}{Q_1} \sum_{i=1}^{Q_1} \varepsilon_i W_{p}^{t+1} \rightarrow W^*,
  \]

where \( \varepsilon_i = \frac{1}{Q_2} \sum_{q=1}^{Q_2} \varepsilon_q \), \( W^* \) is the true global optimal solution.

Theorem 1: Let \( Q_1 = \frac{1}{2} Q_2 + 1 \), and let \( T \) be the total number of iterations. Moreover, let \( c^* = \frac{1}{2} L_x, d^* = \gamma_1 \sqrt{N}, \) where \( \gamma_1 > 1 \) and \( \gamma_2 \geq \frac{1}{\sqrt{1+\alpha}} \frac{c}{d} \).

Then, under Assumptions, the sequence \( \{W^t, H^t\} \) satisfies

\[
\mathbb{E} \left[ \frac{1}{T} \sum_{t=1}^{T} \left( \frac{1}{m} \sum_{i=1}^{m} \|\nabla_v F(W^t, H^t)\|^2 + \frac{\varepsilon_0}{2m} + \frac{\varepsilon_1}{m} \right) \right] \\
\leq \left( \frac{2\sqrt{\varepsilon_0^2 + \varepsilon_1^2}}{m} + \frac{\varepsilon_2}{m} \right) + \frac{c_2}{2}.
\]


Numerical Results II

Application to Item Recommendation:

- Recommendation performance:

References


