20

Deep learning without data aggregation from nodes

Asynchronous consensus algorithm on distributed networks

Abstract

While use of massive data benefits on training DNN models, aggregating all data into one physical location (e.g. a cloud data center) may not be possible due to data privacy concerns from consumers. For example, according to EU GDPR, it is preferable to minimize data transmission between processing nodes.

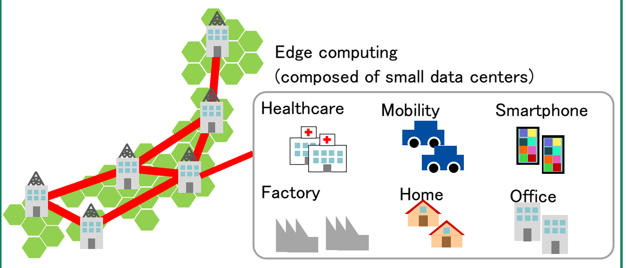
Our goal is to construct training algorithms to obtain a global DNN model that can be adapted to all data, even when individual nodes only have access to different subsets of the data. We assume that this algorithm is allowed to communicate autonomously between nodes, exchanging information such as model variables or their update difference, but data are prohibited from being moved from node they reside on.

Now, several platformers provide advanced services by aggregating/monopolizing data. However, we aim to create a society where data ownership belongs to individual and can be used for a variety of services while protecting data privacy.

Background, goal

Background: We are entering an era of distributed aggregation of due to data volume, privacy protection and legal regulations (e.g. GDPR).

Goal: To obtain a global DNN model without data aggregation (where asynchronous communication among nodes, such as model variable exchange, is allowed).



Problem

Problem: When the data at each node is statistically heterogeneous, a global DNN model cannot be obtained by just minimizing each nod cost function.

Approach: We solve problem by minimizing sum of cost functions under a consensus constraint that all node models are identical with each other.

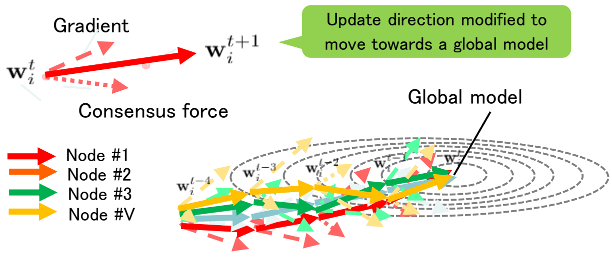
$$\inf_{\{\mathbf{w}_i | i \in \mathcal{V}\}} \sum_{i \in \mathcal{V}} F_i(\mathbf{w}_i; \mathbf{x}_i)$$

$$\text{s.t. } \mathbf{A}_{i|j}\mathbf{w}_i + \mathbf{A}_{j|i}\mathbf{w}_j = \mathbf{0} \qquad \mathbf{A}_{i|j} = \begin{cases} \mathbf{I} & (i > j,\ j \in \mathcal{N}(i)) \\ -\mathbf{I} & (j > i,\ j \in \mathcal{N}(i)) \end{cases}.$$

Data sets are placed across V nodes ($x_1,\cdots,x_V$). Model variables are updated (i) such that minimizes sum of cost functions (Σ Fi) (ii) under a consensus constraint that models are identical among V nodes (s.t....)

Asynchronous consensus algorithm

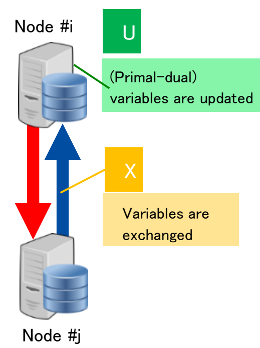
Proposed algorithm: A training algorithm is constructed to obtain a global model by asynchronously exchanging primal model variables and Lagrangian dual variables among nodes. (this enables to work on arbitrary network structure).

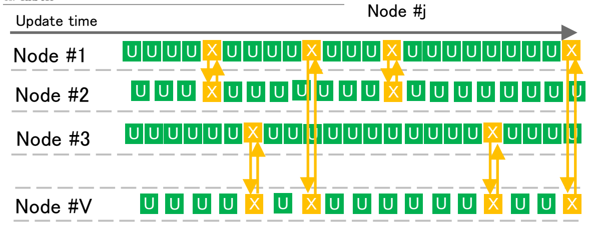


Proposed algorithm

Algorithm 1 PDMM SGD/ADMM SGD

Initialization of $\tilde{\mathbf{z}}_{i|j}^{(0)}, \mathbf{w}_i^{(0)}$
for $k \in \{0, \dots, K-1\}$ **do**
  ▷ Step 1: Update model for each node
  for $i \in \mathcal{V}$ **do**
    $\mathbf{w}_i^{k+1} \leftarrow \Big(\mu \mathbf{w}_i^k - \nabla F_i(\mathbf{w}_i^k; \mathbf{x}_{i,\mathrm{MB}}^k) + \sum_{j \in \mathcal{N}(i)} \big(\alpha \mathbf{A}_{i|j}^{\mathrm{T}} \tilde{\mathbf{z}}_{i|j}^k + \gamma \mathbf{w}_j^k\big)\Big) \Big/ \big(\mu + \alpha|\mathcal{N}(i)| + \gamma|\mathcal{N}(i)|\big),$
    for $j \in \mathcal{N}(i)$ **do**
      $\tilde{\mathbf{y}}_{i|j}^{k+1} \leftarrow \tilde{\mathbf{z}}_{i|j}^k - 2\mathbf{A}_{i|j}\mathbf{w}_i^{k+1}$
    end for
  end for
  ▷ Step 2: Exchange and update variables at random time k
  for $i \in \mathcal{V}$ **do**
    Select $j \in \mathcal{N}(i)$ at random
    Transmit$_{j \to i}(\mathbf{w}_j^{k+1}, \tilde{\mathbf{y}}_{j|i}^{k+1})$
    $\begin{cases} \tilde{\mathbf{z}}_{i|j}^{k+1} \leftarrow \tilde{\mathbf{y}}_{j|i}^{k+1} & \text{(PDMM SGD)} \\ \tilde{\mathbf{z}}_{i|j}^{k+1} \leftarrow \theta \tilde{\mathbf{y}}_{j|i}^{k+1} + (1-\theta)\tilde{\mathbf{z}}_{i|j}^k & \text{(ADMM SGD)} \end{cases}$
  end for
end for





References

[1] T. Sherson, R. Heusdens, and B. Kleijn, Derivation and analysis of the primal-dual method of multipliers based on monotone operator theory, *IEEE transactions on signal and information processing over networks 5*, 2 (2018), 334–347.

[2] K. Niwa, N. Harada, G. Zhang, B. Kleijn, Edge-consensus learning: deep learning on P2P networks with nonhomogeneous data, submitted to KDD 2020.

Contact

Kenta Niwa Email: cs-openhouse-ml@hco.ntt.co.jp

Communication Science Laboratory, Media Intelligence Laboratory

オープンハウス 2020