20 データを端末から漏洩させない分散深層学習

分散NW上で機械学習をするための非同期合意形成技術

どんな研究

現在の深層学習では、1か所に集約したデータを使ってモデルを学習することが一般的です。しかし、データ量の激増やプライバシー保護の観点から、近い将来データは分散蓄積されるようになるでしょう。多端末に分散蓄積されたデータを外に出すことなく、機械学習モデルを最適化する手法を提案します。

どこが凄い

多端末に蓄積されたデータは、統計的に偏っていると仮定することが自然です（例：各端末には一部クラスのデータしか存在しない）。その状況で、端末同士がモデル等の変数を非同期に交換（通信）しながら、全データを使って獲得したかのようなグローバルモデルを学習するアルゴリズムを開発しました。

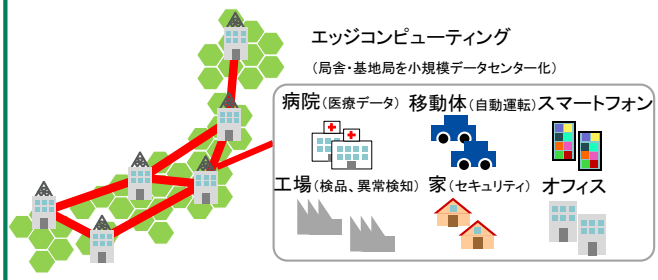
めざす未来

現在は、一部のアプリケーションプラットフォームがデータを集約/独占することにより、高度なサービスを提供しています。データの所有権を個人に帰属させ、プライバシーを保護しながら、多様なサービスに利用できる社会をめざしています。

目的・アプリケーション

背景: データ量、プライバシー保護、法的規制（e.g. GDPR）の観点でデータを分散蓄積する時代になる。

目的: 分散蓄積されたデータを外に出すことなく、深層学習モデルを最適化したい （ただし、モデルなどの変数を非同期で交換（通信）することは許容）。



問題の難しさ

難しいポイント: 各ノードにあるデータが統計的に偏っているとき、各ノードの評価関数を最小化するだけでは、グローバルモデルは得られない。

アプローチ: 全ノードのモデルが一致する制約条件下で、評価関数和を最小化する問題を解く。

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

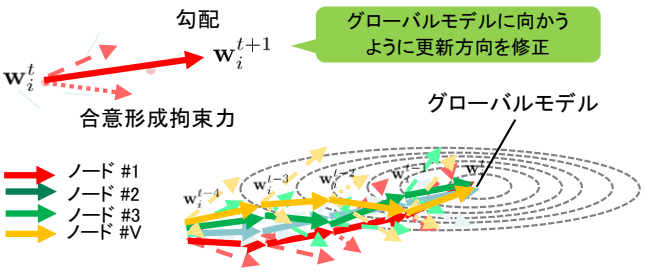
合意形成

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

分散型深層学習: データ（x_i）がVノードに渡って散財蓄積している。
最終的にノード間でモデルが一致する制約下（s.t.…）で、
評価関数和（Σ F_i）を最小化するようにモデル変数（w_i）を更新

非同期分散型深層学習アルゴリズム

提案方式: モデル主変数とラグランジュ双対変数をノード間で非同期に交換しながら、グローバルモデルを得るための学習アルゴリズムを構築（任意のネットワーク構造で動作可）。

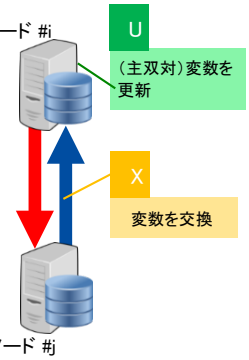


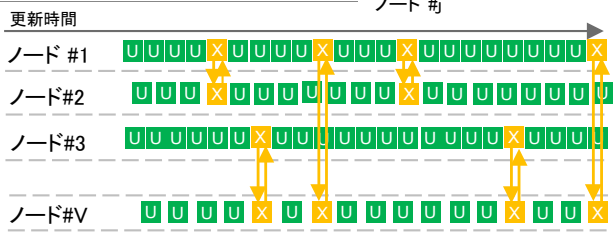
構築したアルゴリズム

Algorithm 1 PDMM SGD/ADMM SGD

1: Initialization of $\hat{\mathbf{z}}_{i|j}^{(0)}, \mathbf{w}_i^{(0)}$
2: **for** $k \in \{0,\dots,K-1\}$ **do**
3: ▷ Step 1: Update model for each node
4: **for** $i \in \mathcal{V}$ **do**
5: $\mathbf{w}_i^{k+1} \leftarrow \left(\rho\mathbf{w}_i^k - \nabla F_i(\mathbf{w}_i^k;\mathbf{x}_{i,\mathrm{MB}}^k) + \sum_{j\in\mathcal{N}(i)}\left(\alpha\mathbf{A}_{i|j}^{\mathrm{T}}\hat{\mathbf{z}}_{i|j}^k + \gamma\mathbf{w}_i^k\right)\right) \Big/ \left(\mu+\alpha|\mathcal{N}(i)|+\gamma|\mathcal{N}(i)|\right)$.
6: **for** $j \in \mathcal{N}(i)$ **do**
7: $\tilde{\mathbf{y}}_{i|j}^{k+1} \leftarrow \hat{\mathbf{z}}_{i|j}^k - 2\mathbf{A}_{i|j}\mathbf{w}_i^{k+1}$
8: **end for**
9: **end for**
10: ▷ Step 2: Exchange and update variables at random time k
11: **for** $i \in \mathcal{V}$ **do**
12: Select $j \in \mathcal{N}(i)$ at random
13: Transmit$_{j\to i}(\mathbf{w}_j^{k+1}, \tilde{\mathbf{y}}_{j|i}^{k+1})$
14: $\begin{cases} \hat{\mathbf{z}}_{i|j}^{k+1} \leftarrow \tilde{\mathbf{y}}_{j|i}^{k+1} & \text{(PDMM SGD)} \\ \hat{\mathbf{z}}_{i|j}^{k+1} \leftarrow \theta\tilde{\mathbf{y}}_{j|i}^{k+1} + (1-\theta)\hat{\mathbf{z}}_{i|j}^k & \text{(ADMM SGD)} \end{cases}$
15: **end for**
16: **end for**

U
X





関連文献

[1] T. Sherson, R. Heusdens, 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*, Vol. 5, 2, pp. 334-347, 2018.
[2] K. Niwa, N. Harada, G. Zhang, B. Kleijn, “Edge-consensus learning: deep learning on P2P networks with nonhomogeneous data,” submitted to KDD 2020

連 絡 先

丹羽 健太（Kenta Niwa） コミュニケーション科学基礎研究所／メディアインテリジェンス研究所
Email: cs-openhouse-ml@hco.ntt.co.jp

オープンハウス 2020