

どんな研究

ネットワーク(NW)で接続された複数の計算機(ノード)を効率活用する**分散モデル学習**は、多くのデータを使って学習するAI技術の根幹です。**ノード間通信やモデル平均化(重みづけ和)**をどのように設計すると、高速かつ高精度に分散モデル学習を実現できるかについて研究を行いました。

どこが凄い

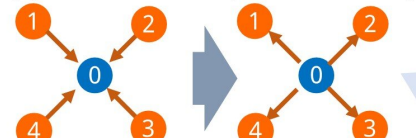
任意のノード数や最大接続数(次数)をもつNWにおいて、分散モデル学習を高速かつ安定させる**(有限時間収束性を持つ)グラフ構成技術**を考案しました。各ノードが統計的に偏ったデータセットを保有する状況で評価したところ、高速かつ安定な学習を実現できることを確認しました。

めざす未来

複数のデータセンタが相互に接続し、データセンタが分散化する未来において、大規模な計算資源を効率的に使った分散モデル学習は不可欠な技術です。**演算や通信の回数を最小化**しながら有限時間収束性を満たす本技術は、**データセンタ全体の消費電力低減**につながるでしょう。

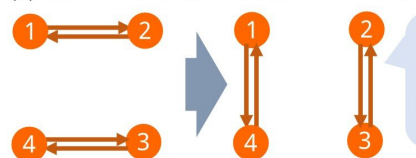
目的・従来方式

- 目的: データセンタに分散した計算資源データを効率的に集めて集合知モデル(e.g. 画像認識や言語処理モデル)を学習したい。
- 有限時間収束グラフ: n 台のノードが次数 k (一時刻辺りの最大接続数)のNWにおいて、各ノードが通信や演算(重み付け加算)を通じて、ローカルモデル $\{x_1, x_2, \dots, x_n\}$ の平均モデル $\bar{x} = \frac{x_1 + \dots + x_n}{n}$ を保有可能なグラフを設計できれば良い。

(a) 中央集権型グラフ (ノード数 n が任意)

n 台のノードから中央サーバにモデルを集約し、平均化したモデルをクライアントに配布することで、有限時間収束を達成。しかし、 n が大きい時、中央サーバの計算/通信負荷が大。

	中央サーバ#0	ノード #1	ノード #2	ノード #3	ノード #4
初期値		x_1	x_2	x_3	x_4
1回目	$\frac{x_1 + x_2 + x_3 + x_4}{4}$				
2回目	$\frac{x_1 + x_2 + x_3 + x_4}{4}$	$\frac{x_1 + x_2 + x_3 + x_4}{4}$	$\frac{x_1 + x_2 + x_3 + x_4}{4}$	$\frac{x_1 + x_2 + x_3 + x_4}{4}$	$\frac{x_1 + x_2 + x_3 + x_4}{4}$

(b) 分散型グラフ (ノード数 n を2のべき乗、次数 k を1に限定)

ノードペア間でモデル平均化を繰り返すことで、有限時間収束を達成。しかし、ノード数 n を2のべき乗に限定することが必要。

	ノード #1	ノード #2	ノード #3	ノード #4
初期値	x_1	x_2	x_3	x_4
1回目	$\frac{x_1 + x_2}{2}$	$\frac{x_1 + x_2}{2}$	$\frac{x_3 + x_4}{2}$	$\frac{x_3 + x_4}{2}$
2回目	$\frac{x_1 + x_2 + x_3 + x_4}{4}$	$\frac{x_1 + x_2 + x_3 + x_4}{4}$	$\frac{x_1 + x_2 + x_3 + x_4}{4}$	$\frac{x_1 + x_2 + x_3 + x_4}{4}$

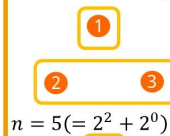
(c) 分散型NW (任意のノード数 n 、任意の次数 k)

任意の n, k に対して有限時間収束を達成する分散型グラフは存在しない。(複数の)データセンタが相互に接続し、多数の計算機ノードを使って分散モデル学習する用途では、大きなノード数 n や任意の次数 k (NWの制限)を扱える有限時間収束グラフが不可欠。

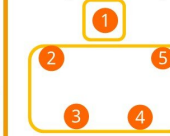
提案方式と実験

- 任意のノード数 n 、任意の次数 k を扱える有限時間収束グラフ(Base- $(k+1)$ グラフ)を提案 [1]。
- メインアイデア: n 個のノードを2のべき乗数のノードで構成された部分集合に分解(左下図)。部分集合内の平均化と部分集合間の情報交換を繰り返すことで、有限時間収束を達成(右下図)。

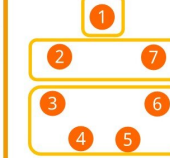
$n = 3 (= 2^1 + 2^0)$



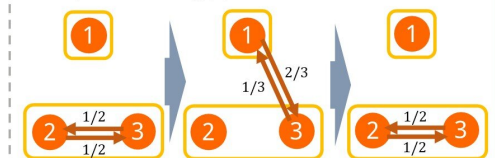
$n = 5 (= 2^2 + 2^0)$



$n = 7 (= 2^2 + 2^1 + 2^0)$



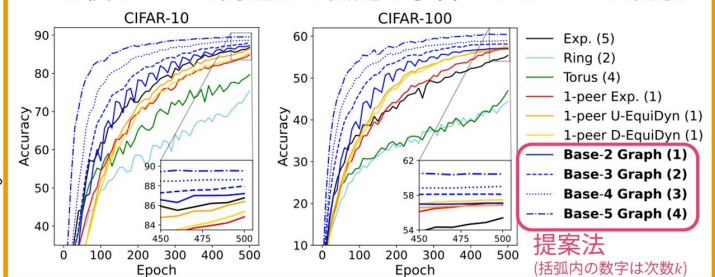
$n = 3, k = 1$ の場合



	ノード #1	ノード #2	ノード #3
初期値	x_1	x_2	x_3
1回目	x_1	$\frac{x_2 + x_3}{2}$	$\frac{x_2 + x_3}{2}$
2回目	$\frac{x_1 + x_2 + x_3}{3}$	$\frac{x_2 + x_3}{2}$	$\frac{2x_1 + 1}{3} \cdot \frac{x_2 + x_3}{2}$
3回目	$\frac{x_1 + x_2 + x_3}{3}$	$\frac{x_1 + x_2 + x_3}{3}$	$\frac{x_1 + x_2 + x_3}{3}$

上記は $n = 3, k = 1$ に限定して例示したが、任意の n, k に対する有限時間収束グラフをアルゴリズムとして表現(ソースコードを[2]に掲載)。

- $n = 25$ ノードが統計的に性質の異なる画像データを保有する状況で、画像認識モデルの分散学習を実施。提案法(Base- $(k+1)$ グラフ)に基づき、可能な限り大きな次数 k を使うことで高速かつ安定に学習できることを確認。



関連文献

[1] Y. Takezawa, R. Sato, H. Bao, K. Niwa, and M. Yamada, "Beyond exponential graph: communication-efficient topologies for decentralized learning via finite-time convergence", in *Proc. The Thirty-seventh Annual Conference on Neural Information Processing Systems (NeurIPS2023)*, 2023.

[2] <https://github.com/yukiTakezawa/BaseGraph>

連絡先

丹羽 健太 (Kenta Niwa) 協創情報研究部 知能創発環境研究グループ