18 Finite-time convergent graph for fast decentralized learning

Communication-efficient learning over data centers

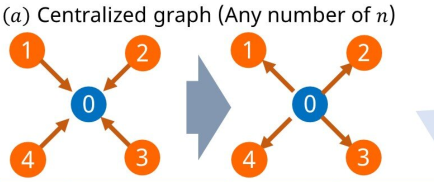
Abstract

Decentralized learning is a fundamental technology to efficiently train machine learning models from a large amount of data by using computing nodes connected over a network (graph). Our proposed Base (k+1) graph guarantees finite-time convergence with any number of nodes and maximum number of connections (degree), enabling fast and stable decentralized learning. We evaluated the efficiency of the graph in a situation where each node has a statistically heterogeneous data subset, and confirmed that it can achieve fast and stable learning of models. This technology, which satisfies finite-time convergence while minimizing the number of operations and communications, will lead to reduce the entire power consumption of data centers.

Goal and related studies

- Goal: To learn collective intelligence models (e.g., image recognition models) by efficiently using computational resource and data on distributed data centers.
- Finite-time convergent graph: In NW with n nodes of degree k (maximum number of connections per a node), it is sufficient to design a graph in which each node can obtain average model of the local model $\{x_1, x_2, \ldots, x_n\}$ as $\bar{x} = (x_1 + \cdots + x_n)/n$ through communication and operations (weighted addition).

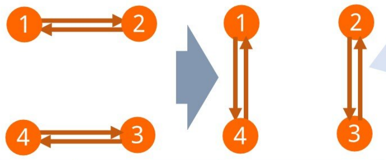
(a) Centralized graph (Any number of n)



Finite-time convergence is achieved by averaging models from n nodes. However, when n is large, computational/communication overheads on the central server are heavy.

	Central server#0	Node #1	Node #2	Node #3	Node #4
Init.		x_1	x_2	x_3	x_4
1st step	$\frac{x_1+x_2+x_3+x_4}{4}$				
2nd step		$\frac{x_1+x_2+x_3+}{4}$	$\frac{x_1+x_2+x_3+x_4}{4}$	$\frac{x_1+x_2+x_3+x_4}{4}$	$\frac{x_1+x_2+x_3+x_4}{4}$

(b) Decentralized graph (Limited to n is a power of two and $k = 1$)



Finite-time convergence is achieved by iteratively averaging models across node pairs. However, it can be achieved only when the number of nodes n is a power of 2.

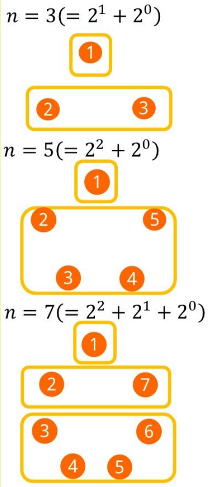
	Node #1	Node #2	Node #3	Node #4
Init.	x_1	x_2	x_3	x_4
1st step	$\frac{x_1+x_2}{2}$	$\frac{x_1+x_2}{2}$	$\frac{x_3+x_4}{2}$	$\frac{x_3+x_4}{2}$
2nd step	$\frac{x_1+x_2+x_3+x_4}{4}$	$\frac{x_1+x_2+x_3+x_4}{4}$	$\frac{x_1+x_2+x_3+x_4}{4}$	$\frac{x_1+x_2+x_3+x_4}{4}$

(c) Decentralized graph (Any number of n, k)

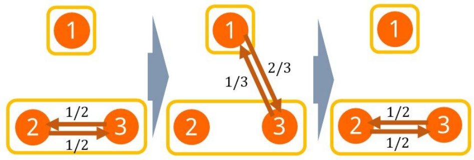
<u>There is no decentralized graph that achieves finite-time convergence for any n and k.</u> For efficiently using a large number of computation nodes interconnecting over multiple data centers for a model training, designing finite-time convergent graphs that can handle a large number of nodes n with any number of degree k is essential.

Proposed method and experiments

- A finite-time convergence graph (Base-(k+1) graph) that can handle any number of nodes n and degree k is proposed [1].
- Main Idea: Decompose n nodes into subsets consisting of a power of 2 nodes (see bottom left figure). Finite-time convergence is achieved by iteratively averaging within subsets and exchanging temporary models between subsets (see bottom right figure).



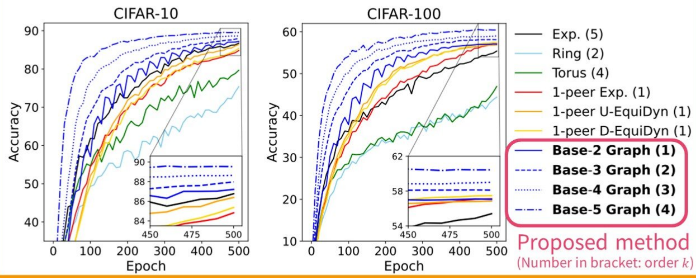
Case: $n = 3, k = 1$



	Node #1	Node #2	Node #3
Init.	x_1	x_2	x_3
1st step	x_1	$\frac{x_2+x_3}{2}$	$\frac{x_2+x_3}{2}$
2nd step	$\frac{x_1+x_2+x_3}{3}$	$\frac{x_2+x_3}{2}$	$\frac{2x_1}{3}+\frac{1}{3}\cdot\frac{x_2+x_3}{2}$
3rd step	$\frac{x_1+x_2+x_3}{3}$	$\frac{x_1+x_2+x_3}{3}$	$\frac{x_1+x_2+x_3}{3}$

Above example is illustrated for n=3 with k=1. Note that, finite-time convergence graphs for any number of n and k can be expressed as algorithms (source code is given in [2]).

- Decentralized learning of image recognition models in a situation where n=25 nodes hold statistically heterogeneous image data subsets was performed. By using proposed graph (Base-(k+1) graph), fast convergence curves were obtained especially when using large order k as possible.



References

[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

Contact

Kenta Niwa, Learning and Intelligent Systems Research Group, Innovative Communication Laboratory