

令和 6 年 6 月 14 日現在

機関番号：62615

研究種目：若手研究

研究期間：2020～2023

課題番号：20K19788

研究課題名（和文）大規模分散深層学習をIn-Network Computingで加速する相互結合網

研究課題名（英文）Interconnection Networks accelerating Large-Scale Distributed Deep Learning with In-Network Computing

研究代表者

河野 隆太（Kawano, Ryuta）

国立情報学研究所・アーキテクチャ科学研究系・特任助教

研究者番号：90855751

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：大規模AIの実用化に向け、深層学習のビッグデータへの適用が喫緊の課題となっている。その解決策として、データセンタ内にドメイン特化型アーキテクチャと呼ばれる演算機構を分散配置し、学習を行うシステムが有望とされているが、従来のデータセンタ向けネットワークでは、遅延などの通信性能の制約により、大規模分散深層学習の高速化が困難である。

本研究では、ネットワーク上で演算の中間処理を行うIn-Network Computingに着目し、低遅延・低頻度の通信と、従来のネットワーク同様の高帯域性・拡張性を両立可能なスイッチ間ネットワークの実用化に向けた諸課題の解決を行った。

研究成果の学術的意義や社会的意義

本研究で提案する相互結合システムでは、高性能計算システム向けに培われた超低遅延ネットワーク技術の1つであるRandom Topologyをデータセンタ向けに取り入れることにより、近年普及の進むIn-Network Computing技術に対し最適化した。すなわち、スイッチ間ネットワークとそれに基づく分散深層学習向けマッピングアルゴリズムにTreeベースの物理・論理トポロジを用いるという従来のシステム設計論を覆し、従来システムと同等以上の高帯域・高拡張性を保証し、既存システムとの親和性を図った。こうした研究成果をOSSとして公開し、提案システムの実用化を推進し、社会需要に応えた。

研究成果の概要（英文）：Toward practical application of large-scale AI, applying deep learning to big data is an urgent issue. As a solution, a system in which computing mechanisms called domain-specific architectures are distributed within data centers is considered to be promising. Conventional data center networks make it difficult to speed up large-scale distributed deep learning due to communication performance constraints such as packet latency. In this research, In-Network Computing, which performs intermediate processing of calculations on the network, is focused on. Various issues are solved in this research, toward the practical application of an inter-switch network that can achieve both low latency and low frequency communication and the same high bandwidth and scalability as conventional networks.

研究分野： 計算機アーキテクチャ

キーワード： 相互結合網 大規模分散深層学習 ビッグデータ In-Network Computing データセンタ

## 様式 C - 19、F - 19 - 1 (共通)

### 1. 研究開始当初の背景

大規模 AI の実用化に向け、深層学習のビッグデータへの適用が喫緊の課題となっている。その解決策として、データセンタ内にドメイン特化型アーキテクチャと呼ばれる演算機構を分散配置し、学習を行うシステムが有望とされているが、従来のデータセンタ向けネットワークでは、遅延などの通信性能の制約により、大規模分散深層学習の高速化が困難である。

### 2. 研究の目的

本研究では、ネットワーク上で演算の中間処理を行う In-Network Computing に着目し、低遅延・低頻度の通信と、従来のネットワーク同様の高帯域性・拡張性を両立可能な、Indirect Random Topology に基づくスイッチ間ネットワークの実用化に向けた諸課題の解決を行った。

### 3. 研究の方法

(1) In-Network Computing に必要な低遅延通信と、低配線コストを両立可能なネットワーク・トポロジの構成手法の開発

スイッチ間ネットワークの配線コストに対し達成可能な遅延性能の上界を求めるための新たな理論を確立した。さらに、この理論に基づき、配線コストに対して通信遅延を最適化可能なネットワークの構成手法 (図 1) を提案した。図に示すように、ネットワークを構成する (番号の振られた) スイッチを (楕円で示した) キャビネットごとにグループ化し、グループ内・グループ外で点対称に配線を構成した。さらに、グループ内の配線数に下限を設けることで、キャビネット内の (低コストな) 配線を増やしつつ、最適化されたキャビネット外配線により通信性能を保証した。

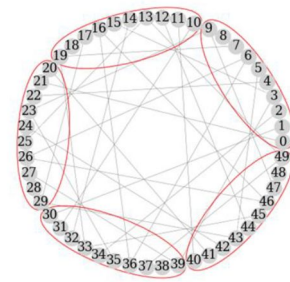
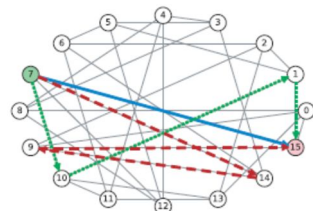
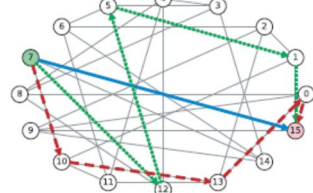


図 1: 低遅延・低配線コストを両立する提案ネットワーク



(a)  $k$ -shortest paths between nodes #7 and #15.



(b)  $k$ -optimized paths between nodes #7 and #15.

図 2: (a) 従来手法で選択可能な複数経路、および (b) 提案手法で選択可能な複数経路

(2) 従来のデータセンタ・ネットワークに必要な高帯域通信を実現するための複数経路を用いたルーティング手法の開発

低直径なスイッチ間ネットワーク上で既存の複数経路ルーティングを用いた場合、ネットワーク上の経路の多様性を活かせず、通信帯域が悪化することが分かっている。

そこで、通信帯域の最大化問題をモデル化した線形プログラムを利用し、最適な複数経路を選択する新たなルーティング手法 (図 2) を探求した。

(3) ドメイン固有アプリケーションの性能向上のためのルーティング手法の開発

実行アプリケーションに対して通信性能を最適化可能なルーティングの動的再構成の理論を構築した。さらに、この理論に基づき、低遅延ネットワークに対し通信の輻輳回避とスケラビリティ向上を両立可能なルーティング手法を開発した。

(4) データセンタ向けネットワークの性能測定の高速度化

ネットワーク構成の最適化において繰り返し実行されるネットワーク性能 (全ノード対最短経路長) の測定を、GPU を

用いて高速に並列計算可能とする新たな手法を開発した。具体的には、データセンタで広く用いられる (総スイッチ数に対して) 次数の低いネットワークを対象とし、ネットワークの全ノード対最短経路長を表す行列の疎行列性に着目した。GPU での並列計算時に行列を List-in-List の疎行列方式で保持することにより、従来手法と比べ時間計算量の大幅な削減を達成した。

(5) 大規模深層学習のためのニューラルアーキテクチャの接続構造最適化

ニューラルネットワークにおけるレイヤ間の接続構造を最適化するため、これまで研究代表者が開発した相互結合網の最適化手法をニューラルネットワーク上に導入し、推論性能の向上と必要計算資源量の削減を両立可能とする手法を開発した。

#### 4. 研究成果

(1) In-Network Computing に必要な低遅延通信と、低配線コストを両立可能なネットワーク・トポロジの構成手法の開発

NPB (NAS Parallel Benchmarks) を用いた評価により、1024 ホスト構成システムにおいて、従来の Fat-Tree を用いたシステムに比べ、ベンチマーク性能が最大 3.12 倍高速化した (図 3)。

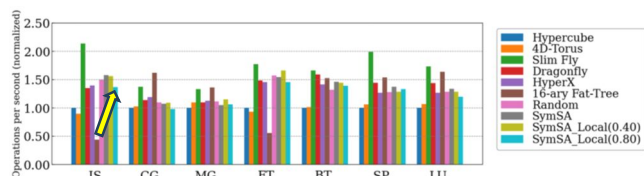


図 3: 提案ネットワークによるベンチマーク性能の向上

(2) 従来のデータセンタ・ネットワークに必要な高帯域通信を実現するための複数経路を用いたルーティング手法の開発

従来の複数経路選択手法である k-shortest path routing に比べ、提案ルーティングは 180 スイッチから構成される Fat-Tree ネットワークでの通信帯域を最大 91.8% 向上させた。さらに、256 スイッチから構成されるランダムネットワークでの通信帯域を最大 18.2% 向上させた。

(3) ドメイン固有アプリケーションの性能向上のためのルーティング手法の開発

64 ノード・次数 4 のスイッチで構成されるランダムネットワークでのベンチマーク性能を最大で 3.61% 向上させた。

(4) データセンタ向けネットワークの性能測定の高速度化

65,536 スイッチから構成されるネットワークにおいて、GPU によるネットワーク性能 (全ノード対最短経路長) の測定を最大で 32.7% 高速化した。

(5) 大規模深層学習のためのニューラルアーキテクチャの接続構造最適化

CIFAR-10 分類タスクを評価対象として、ランダムグラフをニューラルネットワークに用いる従来の手法に比べ、最適化グラフを用いた提案手法が訓練中テスト誤差を最大で 7.18% 改善した。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 KAWANO Ryuta, YASUDO Ryota, MATSUTANI Hiroki, KOIBUCHI Michihiro, AMANO Hideharu	4. 巻 E103.D
2. 論文標題 Traffic-Independent Multi-Path Routing for High-Throughput Data Center Networks	5. 発行年 2020年
3. 雑誌名 IEICE Transactions on Information and Systems	6. 最初と最後の頁 2471 ~ 2479
掲載論文のDOI（デジタルオブジェクト識別子） 10.1587/transinf.2020PAP0005	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計9件（うち招待講演 0件/うち国際学会 3件）

1. 発表者名 Ryuta Kawano
2. 発表標題 Dynamic Routing Reconfiguration for Low-Latency and Deadlock-Free Interconnection Networks
3. 学会等名 10th International Symposium on Computing and Networking (CANDAR'22)（国際学会）
4. 発表年 2022年

1. 発表者名 河野 隆太
2. 発表標題 画像認識のための Optimally Wired Neural Networks
3. 学会等名 第21回情報科学技術フォーラム (FIT 2022)
4. 発表年 2022年

1. 発表者名 河野 隆太
2. 発表標題 経路多様性に着目したニューラルネットワークの配線最適化
3. 学会等名 電子情報通信学会技術研究報告
4. 発表年 2022年

1. 発表者名 Ryuta Kawano
2. 発表標題 GPU Parallelization of All-Pairs-Shortest-Path Algorithm in Low-Degree Unweighted Regular Graph
3. 学会等名 8th ACIS International Virtual Conference on Applied Computing & Information Technology (ACIT 2021) (国際学会)
4. 発表年 2021年

1. 発表者名 河野 隆太
2. 発表標題 低直径ネットワーク・トポロジのための適応型デッドロックフリー・ルーティング
3. 学会等名 電子情報通信学会技術研究報告
4. 発表年 2022年

1. 発表者名 河野 隆太
2. 発表標題 任意トポロジに対する仮想チャネルを用いない低遅延なデッドロックフリー・ルーティング
3. 学会等名 電子情報通信学会技術研究報告
4. 発表年 2021年

1. 発表者名 河野 隆太
2. 発表標題 ルーティングの動的再構成によるネットワークのデッドロックフリー性・低遅延性の両立
3. 学会等名 電子情報通信学会技術研究報告
4. 発表年 2021年

1. 発表者名 Ryuta Kawano
2. 発表標題 Layout-Oriented Low-Diameter Topology for HPC Interconnection Networks
3. 学会等名 8th International Workshop on Computer Systems and Architectures (CSA'20) (国際学会)
4. 発表年 2020年

1. 発表者名 河野 隆太
2. 発表標題 キャビネット内通信を考慮した低直径・配置最適な相互結合網の検討
3. 学会等名 電子情報通信学会 コンピュータシステム研究会 (CPSY) FPGA 応用および一般
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関