様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　４ 年　 ６ 月 １３ 日現在

機関番号：８２４０１
研究種目：若手研究
研究期間：2020～2021
課題番号：２０Ｋ１９８１１

研究課題名（和文）並列I/O最適化による大規模深層学習の高速化に関する研究

研究課題名（英文）Acceleration of large-scale deep learning by optimizing parallel I/O

研究代表者
佐藤　賢斗（Sato, Kento）

国立研究開発法人理化学研究所・計算科学研究センター・チームリーダー

研究者番号：５０７３９６９６

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）：大規模分散深層学習などの大量の学習データ読み込むアプリケーションでは、システムのI/Oの性能が不十分であり、このような深層学習などの新しいアプリケーションに対応するため、I/O性能の重要性が高まっている。このためI/Oの最適化のためにスーパーコンピューター富岳におけるI/O性能の調査、データ圧縮によるI/Oの高速化をおこなった。特に、本研究での知見を活用し深層学習フレームワーク向けソフトウェア開発やMLPerf HPCのベンチマーク評価の一部に貢献した。その結果、「MLPerf HPC」の一つである「CosmoFlow」において「富岳」の約半分の規模を用いた性能評価で世界最高速度を達成した。

研究成果の学術的意義や社会的意義
近年、深層学習に代表される人工知能の研究が盛んに行われており、産業界でも人工知能は様々な形で実用化レベルまで到達している。この深層学習における計算処理には、学習モデルを構築する「学習フェーズ」と、構築された学習モデルを使って、実際に画像認識などの予測・認識を行う「推論フェーズ」に分かれている。深層学習では、より正確な予測・認識を可能にする学習モデルを高速に構築することが重要な要素となっている。本研究は、スーパーコンピュータなどの大規模システムにおける学習フェーズの高速化を達成する研究課題であり、学術的・社会的意義は高いと考える。

研究成果の概要（英文）：Applications that read large amounts of training data, such as large-scale distributed deep learning, have insufficient system I/O performance, thereby, I/O performance is becoming increasingly important to support such applications. To optimize I/O performance, we investigated I/O performance on the supercomputer Fugaku and accelerated I/O by data compression. In particular, our finding from our project partly contributed to the development of software for deep learning frameworks and the benchmark evaluation of MLPerf HPC. As a result, we achieved the world's fastest performance on CosmoFlow, one of the MLPerf HPC benchmarks by using about the half number of Fugaku nodes.

研究分野：高性能計算

キーワード：高性能計算　大規模計算　深層学習　機械学習　I/O　ストレージ

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

近年、深層学習に代表される人工知能の研究が盛んに行われており、産業界でも人工知能は様々な形で実用化レベルまで到達している。この深層学習における計算処理には、学習モデルを構築する「学習フェーズ」と、構築された学習モデルを使って、実際に画像認識などの予測・認識を行う「推論フェーズ」に分かれている。深層学習では、人間の意思決定に近い正確な学習モデルを構築することが重要な要素となっている。このため、より多くの学習データから学習モデルを構築するための学習フェーズの高速化が求められている。

大規模共用計算機における学習フェーズでは、学習データを GFS から読み込む「I/O 処理」と、その学習データから行列演算、畳み込み演算などの「計算処理」からなる。このため、学習フェーズの高速化のためには、「I/O 性能（１秒あたりに読み込む学習データの量）」と「計算性能（１秒あたりに学習するデータ量）」の両方が重要である。

複数の GPU と次世代ストレージによるローカルストレージ(LFS)備えた単一計算ノードもしくは小規模環境のように、I/O 性能がボトルネックとならない環境では、CPU/GPU の性能を効率良く活用することができる。しかし、近年単一計算ノードのメモリでは収まりきらないほど巨大な学習モデルをモデル並列により構築することにより、より正確な学習モデルの構築ができると報告されており、そのためには大規模計算環境においてより多くの学習データを高速に処理する必要がある[1]。しかし、複数のユーザーがバッチスケジューラを介して使用する大規模”共用”計算機では、各計算ノードに搭載されている LFS は計算中のデータの一時保存に限られるため、永続保存が必要なユーザーの学習データは GFS に保存される。そのため、実行時に全ての計算ノードが学習データを読み込むために同時に GFS へアクセスすることとなり、計算ノードあたりの I/O 性能が低くなる。したがって、大規模共用計算機における大規模深層学習では、学習フェーズで使用する計算ノード数(つまり CPU や GPU 数)を増やし計算性能のみを増やしても、並列 I/O 性能がボトルネックとり、学習フェーズの速度がそれ以上向上しない「大規模深層学習のスケール化の問題」が起こる。実際に NVIDIA Tesla P100 GPU を搭載した東京工業大学の大規模共用計算機 TSUBAME3.0（Top500 国内第４位の計算性能）を使用して予備評価を行った。この GPU は、ImageNet の画像データの学習を行うベンチマークにおいて、１秒あたり約 40[MB](= 150 個の画像データ×１画像データあたり 270[KB])の画像データを学習する計算性能を持つ。TSUBAME3.0 はこの GPU を 2160 基搭載しているため、システム全体で 86.4[GB/sec]の画像データの学習が可能な計算性能を持つ。一方、GFS の I/O 性能はシーケンシャルアクセスによる理論性能は 150[GB/sec]であるが、予備評価では深層学習で使用されるランダム読み込みにより 9[GB/sec]の性能しか出ない。このため、TSUBAME3.0 の全ての計算ノードを使って大規模深層学習を行ったとしても 、現状では TSUBAME3.0 性能の約 10%の性能しか引き出せない。2018 年 10 月行われた産業技術総合研究所の大規模共用計算機 ABCI（Top500 国内第１位の計算性能）でのグランドチャレンジでは、ソニーの研究グループが ImageNet を用いた ResNet-50 の学習において、2176GPU を用いて約 3.7 分で完了したと報告[2]があるが、これは一度全ての学習データを GFS から高速な LFS へあらかじめコピーした後に、学習フェーズを実行した場合の実行時間である。GFS からの学習データ読み込みまでを含めた全体の実行時間、つまりユーザーが実際に費やす実行時間はこの時間よりも遥かに長い。

２．研究の目的

本研究の目的は、大規模深層学習のスケール化の問題を解決することである。一般に学習フェーズでは、ミニバッチと呼ばれる複数の画像を結合させたデータセットを生成し、毎回異なるミニバッチを使って反復学習する。このミニバッチの生成の際、局所解に落ちてしまうのを避けるために、複数の画像ファイルをランダムに選び出すこと（シャッフル処理）により、画像データの読み込みがランダムアクセスとなる。これが I/O 性能低下の主な原因となり、GFS の本来の I/O 性能を引き出すことができない。このため、大規模深層学習のための GFS の性能を効率よく引き出すための並列 I/O の最適化をし 10 倍以上の I/O 性能の向上を目的とする。

３．研究の方法

本研究目的の達成のために、まず(1)ミニバッチ生成手法による I/O 性能と学習モデルの精度のトレードオフモデルを構築し、(2)そのモデルに基づいた並列 I/O の最適化システムの開発、(3)GFS を介さないシャッフル処理によるさらなる最適化および(4)検証実験により本研究の目的を達成する。

2020 年度では、まず Chainer などの深層学習フレームワークの学習データ読み込み特性を精緻に明らかにする。本来、シャッフル処理は学習モデルが局所解に落ちてしまうのを避けることが本来の目的である。したがって、最終的な学習モデルの精度が高ければ、必ずしもランダムに

選び出す必要はない。一方、GFS の性能を引き出すためにはシーケンシャル読み込みが最も適しており、いかに GFS に対する読み込みをシーケンシャルにするかが重要となる。そのため、まず（1）GFS の性能を引き出すことができるミニバッチ生成方法と学習モデルの精度のトレードオフモデルを構築する。

その後、この(2)トレードオフモデルに基づき、学習モデルの精度の許容範囲内で、I/O 性能を最大化するための並列 I/O 最適化システムを開発する。しかしながら、この最適 I/O においても、一部にランダムなアクセスが発生することが予想されるが、それを解決するために、実行中に学習データ複製を行い将来の様々なシャッフルパターンを自動生成することにより I/O の高並列化を実現する。必要に応じて、学習データ圧縮や符号化の技術も取り入れる。最終的に、これを新しい学習データフォーマットとして定義し提案並列 I/O システムで複製などの最適化を隠蔽することにより、ユーザーからは単一の学習データ列として見せる。

2021 年度では、シャッフル処理の最適化を行う。民間企業のように、大規模深層学習のための専用機の利用が可能であれば学習データを LFS 上で永続的もしくは長期的に保存することが可能である。一方、国立研究所や大学が所有する多くの大規模計算環境は共用であり複数のユーザが使用するため、LFS は一時保存のみに限られる。そのため実行中にいかに LFS を効率よく使用するかがさらなる最適化に不可欠である。このため、一度 GFS から読まれた学習データは、容量の許す限り各計算ノード上の LFS へ一時保存し、シャッフルの際には(3)GFS を介さずに計算ノード間でシャッフルを行うことによりスケーラブルなシャッフルを実現する。我々は、このシャッフル処理に関して、既に米国フロリダ州立大学と米国ローレンスリバモア国立研究所と共同で、異なるシャッフル処理におけるエントロピーや学習モデルの精度への影響を予備調査[4]しており、ここでの知見を発展させ、どの学習データをどのように GFS から LFS、DRAM、CPU/GPU へ至る階層ストレージ・メモリ上で移動させるべきかを数理最適化問題などに帰着して解決する。その後、最適化された並列 I/O システムと既存の深層学習フレームを統合させ(4)スーパーコンピュータ「富岳」やその他の大規模共用計算機上での検証実験を行う。ここでは、並列 I/O 最適化システムを用いて大規模深層学習を実行した場合と最適化手法を用いない場合の I/O 性能を比べて、10 倍以上高速になっているかを検証し、また全体として大規模深層学習が高速化されていることを確認することにより本研究目標が達成されたか否かを判断する。

４．研究成果

大規模分散深層学習などの大量の学習データ読み込むアプリケーションでは、システムの I/O の性能が不十分であり、このような深層学習などの新しいアプリケーションに対応するため、I/O 性能の重要性が高まっている。このため I/O の最適化のために(1) スーパーコンピューター富岳における I/O 性能の調査、(2) データ圧縮による I/O の高速化をおこなった。

(1)では、大規模分散機械学習のデータ読み込みの高速化に向け、分散機械学習における I/O 性能を富岳上で測定しその結果を分析した。具体的には、第一階層ストレージである LLIO のデータスループットおよびメタデータアクセスの性能測定、ジョブ内の別計算ノードのメモリからのデータ読み込みの性能を行い、富岳における I/O の特性を明らかにした[3,4]。(2)では、深層学習を活用したデータ圧縮ツールである TEZIP の開発を行なった。具体的には、PredNet と呼ばれる Recurrent Neural Network を学習させ、ベースフレームに基づいて将来の画像フレームを予測し、その結果得られる予測フレームと実際のフレームとの差分（デルタフレーム）に対し、量子化などの様々なエンコードを行うことで、高い圧縮率が達成できることを確認した[5]。

(1) (2)での I/O 性能特性調査やデータ圧縮での知見を活用し、国立研究開発法人理化学研究所および富士通株式会社と共同開発した深層学習フレームワーク向けソフトウェア開発や MLPerf HPC のベンチマーク評価の一部に貢献した[6][7]。特に、スーパーコンピュータ規模の処理を必要とする大規模機械学習処理のベンチマーク「MLPerf HPC」の一つである「CosmoFlow」において、スーパーコンピュータ「富岳」の約半分の規模を用いて計測した結果、世界最高速度を達成し第 1 位を獲得した[8]。

[1] Yanping Huang et al., “GPipe: Efficient Training of Giant Neural Networks using Pipeline Parallelism”, 2018, arXiv:1811.06965

[2] Hiroaki Mikami et al. “ImageNet/ResNet-50 Training in 224 Seconds”, 2018, arXiv:1811.05233

[3] Takaaki Fukai, Kento Sato, “Measurement of I/O performance for distributed deep neural networks on Fugaku”, The 3rd R-CCS International Symposium, Feb, 2021

[4] Takaki Fukai, Kento Sato “Measurement of I/O Performance on a Hierarchical File System for Distributed Deep Neural Network”, the 4th R-CCS International Symposium (RCCS-IS4), Kobe, Japan, Feb. 2022 (Lightning Presentation)

[5] Rupak Roy, Kento Sato, Subhadeep Bhattacharya, Xingang Fang, Yasumasa Joti, Takaki Hatsui, Toshiyuki Hiraki, Jian Guo and Weikuan Yu, “Compression of Time Evolutionary Image Data through Predictive Deep Neural Networks”, In the proceedings of the 21 IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGrid 2021), May, 2021

[6] Akihiro Tabuchi, Koichi Shirahata, Masafumi Yamazaki, Akihiko Kasagi, Takumi Honda, Kouji Kurihara, Kentaro Kawakami, Tsuguchika Tabaru, Naoto Fukumoto, Akiyoshi Kuroda, Takaaki Fukai and Kento Sato, “The 16,384-node Parallelism of 3D-CNN Training on An Arm CPU based Supercomputer”, 28th IEEE International Conference on High Performance Computing, Data, and Analytics (HiPC2021), Nov, 2021
[7] Steven Farrell, Murali Emani, Jacob Balma, Lukas Drescher, Aleksandr Drozd, Andreas Fink, Geoffrey Fox, David Kanter, Thorsten Kurth, Peter Mattson, Dawei Mu, Amit Ruhela, Kento Sato,,Koichi Shirahata, Tsuguchika Tabaru, Aristeidis Tsaris, Jan Balewski, Ben Cumming, Takumi Danjo, Jens Domke, Takaaki Fukai, Naoto Fukumoto, Tatsuya Fukushi, Balazs Gerofi, Takumi Honda, Toshiyuki Imamura, Akihiko Kasagi, Kentaro Kawakami, Shuhei Kudo, Akiyoshi Kuroda, Maxime Martinasso, Satoshi Matsuoka, Kazuki Minami, Prabhat Ram, Takashi Sawada, Mallikarjun Shankar, Tom St. John, Akihiro Tabuchi, Venkatram Vishwanath, Mohamed Wahib, Masafumi Yamazaki, Junqi Yin and Henrique Mendonca, “MLPerf HPC: A Holistic Benchmark Suite for Scientific Machine Learning on HPC Systems”, The Workshop on Machine Learning in High Performance Computing Environments (MLHPC) 2021 in conjunction with SC21, Nov, 2021
[8] **スーパーコンピュータ「富岳」が機械学習処理ベンチマーク** MLPerf HPC **で世界第 1 位を獲得 − 深層学習モデル** CosmoFlow **の単位時間あたりの学習で世界最高速度を達成**, https://pr.fujitsu.com/jp/news/2021/11/18-1.html

５．主な発表論文等

〔雑誌論文〕　計3件（うち査読付論文　3件／うち国際共著　2件／うちオープンアクセス　0件）

１．著者名 Rupak Roy, Kento Sato, Subhadeep Bhattacharya, Xingang Fang, Yasumasa Joti, Takaki Hatsui, Toshiyuki Hiraki, Jian Guo and Weikuan Yu	４．巻 -
２．論文標題 Compression of Time Evolutionary Image Data through Predictive Deep Neural Networks	５．発行年 2021年
３．雑誌名 21th IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGRID)	６．最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

１．著者名 Tabuchi Akihiro、Shirahata Koichi、Yamazaki Masafumi、Kasagi Akihiko、Honda Takumi、Kurihara Kouji、Kawakami Kentaro、Tabaru Tsuguchika、Fukumoto Naoto、Kuroda Akiyoshi、Fukai Takaaki、Sato Kento	４．巻 -
２．論文標題 The 16,384-node Parallelism of 3D-CNN Training on An Arm CPU based Supercomputer	５．発行年 2021年
３．雑誌名 2021 IEEE 28th International Conference on High Performance Computing, Data, and Analytics (HiPC)	６．最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/HiPC53243.2021.00029	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Farrell Steven、Emani Murali、Balma Jacob、Drescher Lukas、Drozd Aleksandr、Fink Andreas、Fox Geoffrey、Kanter David、Kurth Thorsten、Mattson Peter、Mu Dawei、Ruhela Amit、Sato Kento et al.	４．巻 -
２．論文標題 MLPerf HPC: A Holistic Benchmark Suite for Scientific Machine Learning on HPC Systems	５．発行年 2021年
３．雑誌名 2021 IEEE/ACM Workshop on Machine Learning in High Performance Computing Environments (MLHPC)	６．最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/MLHPC54614.2021.00009	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

〔学会発表〕　計1件（うち招待講演　0件／うち国際学会　0件）

１．発表者名 Takaaki Fukai, Kento Sato
２．発表標題 Measurement of I/O performance for distributed deep neural networks on Fugaku
３．学会等名 The 3rd R-CCS International Symposium
４．発表年 2021年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

Compression of Time Evolutionary Image Data ... 略
https://www.hpbd.r-ccs.riken.jp/hpbd/en/research/
HPC and AI Initiatives for Supercomputer Fugaku略
https://www.fujitsu.com/global/documents/about/resources/publications/technicalreview/2020-03/article09.pdf
富岳における深層学習フレームワーク構築・最適化とMLPerf HPC ベンチマーク
https://www.riken.jp/pr/news/2020/20201119_1/index.html

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			
米国	Florida States University			