

令和 5 年 6 月 21 日現在

機関番号：13904

研究種目：基盤研究(C)（一般）

研究期間：2019～2022

課題番号：19K11874

研究課題名（和文）深層学習と経験的手法の協調によるメモリアクセス最適化プログラムの自動合成

研究課題名（英文）Automatic synthesis of memory access optimization programs synergistically coordinated with deep learning and empirical methods

研究代表者

佐藤 幸紀 (Sato, Yukinori)

豊橋技術科学大学・工学（系）研究科（研究院）・准教授

研究者番号：30452113

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：メモリアクセス最適化プログラムの自動合成を目指して、深層学習と経験的手法を協調させつつ最適化を適応することに取り組んだ。深層学習に関しては、実行コードレベルの最適化と親和性が高いグラフニューラルネットワークを活用した最適化の可能性の調査やプロトタイプ実装の実施を行い、今後の研究の発展につながる知見を得ることができた。メモリ階層を考慮した性能モデルとしてPolyhedralモデルやキャッシュシミュレータを活用した経験的手法による最適化に関しては、実践的な性能解析ツールやメモリ信頼性判定シミュレータの基盤技術として応用可能であるという知見が得られた。

研究成果の学術的意義や社会的意義

近年の深層学習技術の進展は目覚ましい反面、これらの成果を社会に幅広く展開するためには性能・電力・コストの面で最適化された深層学習処理向けのシステムアーキテクチャを探索していくことが急務である。このような状況の下、ハード・ソフトにまたがる広大なアーキテクチャ設計空間に対して最適化を行う機構の自動合成を目指すのが本研究の目的であった。プログラマによる手作業では手に負えなくなる規模の複雑かつ明示的で緻密な場合分けが必要となる最適化であっても、経験的手法を性能モデルと連携する形で拡張させていくことにより自動化できる可能性が示された。

研究成果の概要（英文）：Toward automatic synthesis of memory access optimization programs, we tackled with synergistically coordinated with deep learning and empirical methods. For the deep learning part, we investigated optimization methods that exploit graph neural networks to smoothly cooperate with code optimization techniques for executable code. From the results of prototype implementation, we obtained some insights for further research. For the empirical method part, we investigated on the Polyhedral model and a cache simulator as a performance model against systems with complex memory hierarchy. From the results of evaluation, we demonstrated that our techniques developed in this work can be applied to a productive performance profiling tool and a simulator that assesses the memory reliability of systems.

研究分野：計算機システム

キーワード：メモリアクセス最適化 性能モデル ハード・ソフト協調設計 深層学習と経験的手法 プログラム自動合成

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

1. 研究開始当初の背景

近年の深層学習技術の進展は目覚ましい反面、これらの成果を社会に幅広く展開するためには性能・電力・コストの面で最適化された深層学習処理向けのシステムアーキテクチャを探求していくことが急務である。このような状況の下、ハード・ソフトにまたがる広大なアーキテクチャ設計空間に対して深層学習を活用し、既存の高位最適化コンパイラに実装されている経験的手法と協調しながら動作する最適化器の自動合成が注目を集めていた。いわば、深層学習の手法を使って深層学習を高速化するシステムアーキテクチャを探求するという試みであり、実環境での観測に基づく大量の教師学習データを得ることにより様々な最適化が自動化できるのではないかという仮説に対して、システムアーキテクチャの設計レベルで評価できる技術もそろいつつあるという状況であった。

2. 研究の目的

深層学習処理向けに特化したシステムアーキテクチャにおいて、個々のソフト・ハードの特徴に由来する性能・電力の面での挙動を忠実に目的関数に反映させることは、これまで行われてきた明示的で緻密な場合分けでの対応は手に負えない規模の複雑なプログラミングを要求する。このような複雑なプログラミング作業は、実環境での観測に基づく大量の教師学習データによりデータ主導型で実現するべきものとの着想の下、Polyhedral 最適化の一部のパスに DNN 学習を応用し、元来から実装されている経験的手法により構築されたパスと協調しながらプログラムをメモリアクセスの観点で最適なものに変換するプログラムの自動合成の実現を目指すことを研究の目的とした。

3. 研究の方法

最適化の一部のパスに DNN 学習と経験的手法を協調させつつ適応するという目的を実現するために、1) DNN 学習や実行コードレベルの最適化と親和性が高いグラフニューラルネットワークを活用した最適化の可能性の調査やプロトタイプ実装の実施、2) メモリ階層を考慮した性能モデルとして Polyhedral モデルやキャッシュシミュレータを活用した最適化手法の高度化と実環境での評価、という2つの項目に特に注力して研究を進めることとした。

1つ目の DNN 学習やグラフニューラルネットワークを活用した最適化に関しては、研究の当初においては多層パーセプトロンを用いて DNN 学習させていく計画であったが、グラフニューラルネットワークの利用が主にソースコードのバグ検出やソフトウェア工学の分野で利用されている事例が報告されていることから評価を行った。

2つ目の Polyhedral モデルやキャッシュシミュレータを活用した最適化については、研究の当初においては Polyhedral 最適化のフローにおいてコード変換の核となる ISL ライブラリにおいてループ反復空間の依存関係や計算コストを表す数学的モデルである SCoP 形式に如何にプログラムが変換されていくかという過程を探り、最適化アルゴリズムの高度化の余地を探求した。探求の結果、Polyhedral モデルの最適化に資するメモリ階層を考慮した性能モデルが有用であろうとの知見を得て、メモリトレースを用いた詳細なキャッシュシミュレータの構築と高度化を進めた。プログラム実行時に透過的かつ高速にメモリトレースを取得するために、バイナリ計装ツールの DynamoRIO に着目し、内製のオンラインキャッシュシミュレータと融合させた Exana-C2Sim というツールを開発し、評価を進めた。

4. 研究成果

1つ目の研究目標である DNN 学習やグラフニューラルネットワークを活用した最適化に関しては、プロトタイプ実装を行い、有用性の検証を行ったが、既存の経験的手法に基づくプログラム最適化技術と同等、あるいは、凌駕する方式を見つけることはできなかった。深層学習やグラフニューラルネットワークで学習した結果として何を出力するか設計が研究開始時に想定していた以上に難しく、単に性能の改善の余地の有無を出力するのか、性能の面で問題となる個所をプログラム中から検出することを出力するのか、あるいは、プログラムのコード自体を書き換えて出力するのが良いであろうかなど、具体的な学習手法に実装する形で実現するためには、学習手法の進展をキャッチしつつ、性能最適化分野の更なる技術の蓄積が必要であるという知見を得た。

2 つ目の研究目標である Polyhedral モデルやキャッシュシミュレータを活用した最適化に関しては、主に 2 つの成果を上げることができた。1 つ目の成果としては、スーパーコンピュータ富岳の CPU である A64FX に搭載されるベクトル命令拡張である SVE に対応した性能解析ツールのプロトタイプ実装を完成させた成果を国際学会で発表し、WANC Best Paper Award を受賞するとともに、スーパーコンピュータ富岳の実際の性能最適化の現場での実利用を見据えた研究に発展させている。2 つ目の成果としては、不揮発性メモリをメモリ階層に組み入れた場合のメモリアクセスの挙動をシミュレーションする統合型メモリ階層シミュレータとして高度化することに成功したことである。DRAM よりも高速かつ低エネルギーの書き込みが可能な VC-MRAM (Voltage-Controlled Magnetoresistive RAM : 電圧駆動磁気抵抗メモリ) をメインメモリにするシナリオにおいて、エラーが確率的に発生するという条件の下で信頼性のモデルを構築し、ECC による誤り訂正や Row バッファレベルでの局所性、ビットレベルでの誤り確率に基づくエラーの挿入に関する詳細なモデリングを行い、システムレベルの信頼性を評価することができた。本研究の成果は国際会議ワークショップで発表され、CSA Best Paper Award を受賞した。また、The 6th cross-disciplinary Workshop on Computing Systems, Infrastructures, and Programming (xSIG 2022) で発表した際には、Outstanding Effort Award を受賞した。

5. 主な発表論文等

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

1. 著者名 Saito Daiki, Hirofuchi Takahiro, Arai Hiroko, Sato Yukinori	4. 巻 -
2. 論文標題 A simulation of a memory subsystem using a highly energy-efficient but erroneous MRAM	5. 発行年 2021年
3. 雑誌名 In Proceedings of 9th International Workshop on Computer Systems and Architectures (CSA 2021)	6. 最初と最後の頁 120-126
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/CANDARW53999.2021.00027	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 Chugo Kazuki, Sato Yukinori	4. 巻 -
2. 論文標題 Thread-Aware Cache Simulator for HPC Application Tuning	5. 発行年 2021年
3. 雑誌名 In Proceedings of 12th International Workshop on Advances in Networking and Computing (WANC 2021)	6. 最初と最後の頁 424-428
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/CANDARW53999.2021.00078	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 -

〔学会発表〕 計4件（うち招待講演 0件 / うち国際学会 3件）

1. 発表者名 Daiki Saito, Takahiro Hirofuchi, Hiroko Arai, Yukinori Sato
2. 発表標題 A simulation of a memory subsystem using a highly energy-efficient but erroneous MRAM
3. 学会等名 9th International Workshop on Computer Systems and Architectures (CSA 2021) (国際学会)
4. 発表年 2021年

1. 発表者名 Kazuki Chugo, Yukinori Sato
2. 発表標題 Thread-Aware Cache Simulator for HPC Application Tuning
3. 学会等名 12th International Workshop on Advances in Networking and Computing (WANC 2021) (国際学会)
4. 発表年 2021年

1. 発表者名 Kazuki Chugo, Yukinori Sato
2. 発表標題 An Online Trace-Driven Cache Simulator for ARM-Based Supercomputers
3. 学会等名 Poster presentation, IEEE Symposium on Low-Power and High-Speed Chips and Systems (COOL Chips 24) (国際学会)
4. 発表年 2021年

1. 発表者名 斉藤 大貴, 広淵 崇宏, 荒井 礼子, 佐藤 幸紀
2. 発表標題 ペリファイ機構を搭載したVC-MRAMに対するシステムレベルシミュレーション
3. 学会等名 The 6th cross-disciplinary Workshop on Computing Systems, Infrastructures, and Programming (xSIG 2022)
4. 発表年 2022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