

令和元年9月11日現在

機関番号：10101

研究種目：基盤研究(B) (一般)

研究期間：2015～2018

課題番号：15H02673

研究課題名(和文) データマイニングを加速する次世代リコンフィギュラブルアーキテクチャの創出

研究課題名(英文) Reconfigurable Architectures for Accelerating Data Mining

研究代表者

本村 真人 (Motomura, Masato)

北海道大学・情報科学研究科・教授

研究者番号：90574286

交付決定額(研究期間全体)：(直接経費) 13,800,000円

研究成果の概要(和文)：本研究では組合せ最適化問題の最適解を導き出すことを狙ったアーキテクチャ研究に注力した。従来の単一スパースハードウェアグラフでは、ハードウェア構造より密なグラフを再現する際、スピンを複製し、擬似的に密なグラフを再現するマイナーエンベディングという方法がとられる。スピンの状態の更新が、隣接スピンの状態とスピン間相互作用の積和演算によって再現されることを利用して、スパースなハードウェア構造それぞれに対して、時間方向の接続(積和演算を継続)を可能にすることで、複製スピンによる強い相互作用を加えることなく処理を行うことを実現し、大規模な問題から高い精度の解を得るアーキテクチャを提案した。

研究成果の学術的意義や社会的意義

IoT 社会の到来により、データマイニングに代表されるビッグデータ処理が計算処理の中心的課題となりつつある。本研究は、(1)データマイニング処理に適したリコンフィギュラブルアーキテクチャと、(2)HW アーキテクチャを考慮したデータマイニングアルゴリズムの二つの課題に統括的に取り組むことで、超高速・低電力なデータマイニング処理基盤の確立を目指すものである。

研究成果の概要(英文)：We have conducted researches of architectures for combinatorial optimization problems. Conventional idea has been to conduct minor-embedding or a target graph onto hardware graph that is more sparse than the former. Our new architectural proposal is to time-multiplexing a hardware structure for expanding hardware graph, that is powerful enough to sustain the original denser target graph. This idea has been verified to achieve higher accuracy for large graphs in our simulation.

研究分野：集積回路、計算機アーキテクチャ

キーワード：ビッグデータ データマイニング イジングモデル 組合せ最適化問題

様式 C-19、F-19-1、Z-19、CK-19 (共通)

1. 研究開始当初の背景

IoT(Internet of Things)技術の進展とともに、身の回りに溢れる大量の小型端末(各種センサ、モバイル端末、IC タグ等)が生成するビッグデータの中から人間社会にとって有益な情報をどのようにして取り出すかが、現代社会の学術的重要課題として浮上している。この分野は、ストリームデータにおける特異値や変化点等をリアルタイム検出することに着目した場合には CEP(Complex Event Processing)、頻出パターンや分類規則などの有用な知識を発掘するような深い情報処理を行う場合にはデータマイニング、更にストリームからのリアルタイム知識発掘に着目する場合にはストリームマイニング等と呼ばれている。

データマイニング(ここでは CEP やストリームマイニングも含み広義にこう呼ぶ)はデータセンターにおけるビッグデータ処理の根幹技術として既に注目されていたが、今後小型端末が爆発的に増殖するに従い、ネットワークに送出する手前でデータトラフィックを減らす処理(=小型端末のスマート化)としても更に重要性を増すと予想された。更なる IoT 利活用に際し、データセンターのエネルギー消費量削減および小型端末の環境負荷低減は社会的に大きな課題であり、超高速・低電力なデータマイニング処理の実現を目指す研究の学術的価値は極めて高いと考えられた。

2. 研究の目的

《目的 1: リンコンフィギュラブルアーキテクチャとデータマイニングアルゴリズムの共創》

過去数十年に亘り、例えば文字列検索やパターン照合・距離計算などを目的として新たな専用アーキテクチャが種々提案されてきた(例えば辞書検索プロセッサ[Motomura, ISSCC90])。これらは単純並列演算による処理加速を基本的な特徴としており、計算量・使用メモリ量を賢く減らすことを主眼とするアルゴリズム技術の発展方向とは得てして相いれない(従って実用に供され難い)という欠点があった。本研究では、アーキテクチャとアルゴリズムの専門家による共同研究を通じ、データマイニングを加速する統括的な技術基盤の構築を目指した。

《目的 2: FPGA を代替する新たなリンコンフィギュラブルアーキテクチャの創出》

本研究は、広義のデータマイニングが今後のコンピューティングの中心的処理課題であるとの認識のもと、そのアルゴリズム要求にアーキテクチャが応えることが、次世代のメインストリーム・リンコンフィギュラブルアーキテクチャの創出につながるという立場に立脚している。本研究の成果は、小型端末側では LSI 内部に搭載する主にストリームマイニング向けリンコンフィギュラブル IP コアとして、データセンター側では FPGA を代替する新たなビッグデータ処理アクセラレータとして、これらの処理の高速化・低電力化に大きな意味を持つ。

3. 研究の方法

前半は、データマイニングの中でも難易度の高い頻出アイテムセットマイニングのアーキテクチャ研究に取り組む。エッジ側で高頻度に生成されるデータに対して、従来の CPU 上のソフトウェアによる処理では、OS を含む多数のソフトウェアスタックを経由することにより、処理遅延の増加とそれに伴う性能低下という問題が存在する。リアルタイムなエッジデバイス制御 には、ソフトウェアスタックを介さない、低遅延な処理方式が求められている。そこで、頻出アイテムセットマイニング処理アルゴリズムをハードウェア化し、FPGA を搭載するネットワーク・インターフェース上に実装することで、遅延に関する問題を解決する。ネットワークを通して得たデータを、CPU を介することなく FPGA 上の専用回路によりマイニングすることで、低遅延に処理を行うことが可能になる。ただし、従来のデータを蓄積してから処理を行うマイニングアルゴリズムを単純に FPGA 上の専用回路化するだけでは、結果が得られるまでの遅延を短縮することは困難である。そこで、データが到着する毎にインクリメンタル処理を行うストリーム処理型頻出アイテムセットマイニング方式とそのハードウェアアーキテクチャを研究する。

4. 研究成果

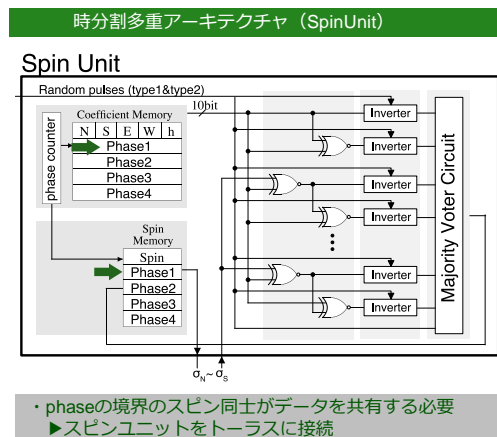
FPGA 実験の結果、ベースとなるソフトウェア実装の Skip LC-SS アルゴリズム[Yamamoto, Sigmod'14]と比較して、約 100 倍の速度改善と 100 倍の省メモリ化を達成した。これにより、ネットワーク・インターフェース上に FPGA を配置し、ソフトウェア処理をスキップすることで、従来の計算機よりもリアルタイムにデータの観測と集約処理が可能であるという知見を得た。

一方、ビッグデータ時代や Society5.0 の時代に重要となる組合せ最適化問題を効率よく解くリンコンフィギュラブルアーキテクチャに関して、特に大きな成果を上げることができた。組合せ最適化問題向けの非ノイマン型計算アーキテクチャとして、イジングモデルの基底状態を探索するアニーリング計算機がある。イジングモデルの基底状態探索問題は、組合せ最適化問題の最小エネルギーの探索問題に置き換えることで、組合せ最適化問題を解くことが可能となる。しかし、現状の FPGA ベースのアニーリング計算機では、搭載可能なスピン数が少ないため、実用的な組合せ最適化問題に適用することは難しい。本研究は、大規模な組み合わせ最適化問題を解くために、対象の問題を時間方向に分割し FPGA のオンチップメモリを活用するイジング計算機を提案する。結果として、従来アニーリング計算機と比較して、約 64 倍のスピンが搭載可能であるということがわかった。

時分割多重処理を実現する提案アーキテクチャでは、すべてのスピンの状態を記憶するスピンメモリ(Spin Memory)と隣接スピン間の相互作用係数を記憶する相互作用係数メモリ(Coefficient Memory)、及び、一列ずつスピンの状態の更新を行う Operator / Pipeline Register(PR) アレイによって構成される Spin Unit で構成される(下図)。スピンメモリは、スピンをキメラトポロジーのローカル

スピンをまとめて保存する。Spin Update Unit では、隣接するスピン同士は同時更新できないという制約を破らないため、Operator の間にパイプラインレジスタを挟むような構成になっている。さらにパイプライン処理により、メモリの読み出し・書き込み分のレイテンシによるスループットの低下を防いでいる。具体的な動作としては、キメラトポロジーの Phase1 のローカルスピン 0 を読み出し、Spin update unit を転送する。次のタイミングで、Phase2 のローカルスピン 0 を Spin Memory から、Phase0 の隣接スピンの相互作用係数を Coefficient Memory から取得し Spin Update Unit へ転送する。Spin Update Unit は取得した値をもとに、Spin Update Unit で Phase1 のローカルスピン 0 に関する処理を行う。前段に PR が挿入されている Operator に関しては、挿入されていないオペレータから更新済みのスピンの情報をフォワーディングする。これを同様に Phase4 までパイプライン処理を行う。その後、Phase1 に戻り、ローカルスピン 1 に関する処理を行い、すべてのローカルスピンを更新するまで繰り返す。これを 1 回の更新とし、イジングモデルが収束するまでこれを繰り返す。

従来の単一スパースハードウェアグラフでは、ハードウェア構造より密なグラフを再現する際、スピンを複製し、擬似的に密なグラフを再現するマイナーエンベディングという方法がとられていた。本提案では、スピンの状態の更新が、隣接スピンの状態とスピン間相互作用の積和演算によって再現されることを利用して、スパースなハードウェア構造それぞれに対して、時間方向の接続(積和演算を継続)を可能にすることで、複製スピンによる強い相互作用を加えることなく処理を行うことを実現し、大規模な問題から高い精度の解を得るアーキテクチャを実現できる。



5. 主な発表論文等

[雑誌論文](計 3 件)

1. Tanibata A., Schmid A., Takamaeda-Yamazaki S., Ikebe M., Motomura M., and Asai T., "Proto-computing architecture over a digital medium aiming at real-time video processing," Complexity, vol. 2018, pp. 3618621-1-11 (2018).
2. Hida I., Takamaeda-Yamazaki S., Ikebe M., Motomura M., and Asai T., "A high performance and energy efficient microprocessor with a novel restricted dynamically reconfigurable accelerator," Circuits and Systems, vol. 8, no. 5, pp. 134-147 (2017).
3. Yamamoto K., Ikebe M., Asai T., and Motomura M., "FPGA-based stream processing for frequent itemset mining with incremental multiple hashes," Circuits and Systems, vol. 7, no. 10, pp. 3299-3309 (2016).

[学会発表](計 6 件)

1. Tanibata A., Schmid A., Takamaeda-Yamazaki S., Ikebe M., Motomura M., and Asai T., "FPGA implementation of edge-guided pattern generation for motion-vector estimation of textureless objects (demo night)," The 27th International Conference on Field-Programmable Logic and Applications, Culture and Convention Center Het Pand, Ghent, Belgium (Sep. 4-8, 2017).
2. Yamamoto K., Ikebe M., Asai T., Motomura M., and Takamaeda-Yamazaki S., "Time-Division Multiplexing," GI-CoRE GSQ, GSB, & IGM Joint Symposium -Quantum, Informatics, Biology, & Medicine -, Hokkaido University, Sapporo, Japan (Jul. 10-11, 2017).
3. Tanibata A., Ushida M., Schmid A., Ikebe M., Asai T., and Motomura M., "A hardware cellular-automaton architecture for spatial pattern generation towards motion-vector estimation of textureless objects," 2016 International Symposium on Nonlinear Theory and its Applications, pp. 622-625, New Welcity Yugawara, Shizuoka, Japan (Nov. 27-30, 2016).
4. Yamamoto K., Huang W., Takamaeda-Yamazaki S., Ikebe M., Asai T., and Motomura M., "A Time-Division Multiplexing Ising Machine on FPGAs," International Symposium on Highly-Efficient Accelerators and Reconfigurable Technologies (HEART 2017), Ruhr University, Bochum, Germany (Jun. 7-9, 2017).
5. Hida I., Ikebe M., Asai T., and Motomura M., "A two-clock-cycle naive Bayes classifier for dynamic branch prediction in pipelined RISC microprocessors," 2016 IEEE Asia Pacific Conference on Circuits and Systems, Ramada Plaza Jeju Hotel, Jeju, Korea (Oct. 25-30, 2016).
6. Yamamoto K., Asai T., and Motomura M., "Hardware architecture for online frequent

items mining with memory-efficient data structure," COOL Chips XIX, Yokohama Media & Communications Center, Yokohama, Japan (Apr. 20-22, 2016).

〔図書〕(計 1 件)

本村 真人、「FPGA の原理と構成」、天野 英晴 編、株式会社オーム社 (2016). [第 8 章を執筆]

〔産業財産権〕

○出願状況(計 0 件)

〔その他〕

ホームページ等

<http://lalsie.ist.hokudai.ac.jp/jp/>

6. 研究組織

(1)研究分担者

研究分担者氏名：有村 博紀

ローマ字氏名：Arimura, Hiroki

所属研究機関名：北海道大学

部局名：大学院情報科学研究科

職名：教授

研究者番号(8 桁)：20222763

※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、研究者個人に帰属されます。