

# GPU および FPGA を活用したビッグデータ分析の高速化

## Accelerating big data analysis using GPU and FPGA

天笠俊之

筑波大学計算科学研究センター

### 1. 研究目的

ビッグデータ処理において、GPU や FPGA による高速化が注目されている。GPU および FPGA はそれぞれ異なる特性を持っており、GPU は高いメモリバンド幅による高スループットな処理が可能な反面、レイテンシは高い。一方で FPGA は低レイテンシ処理が可能な反面、スループットでは GPU に劣る。このため、特定のワークロードに対して、これらを適切に使い分ける、あるいは両者を組み合わせることで、性能をさらに引き出すことが期待できる。すなわち大規模な祖粒度並列処理部分は従来通り GPU が担当しつつ、GPU では性能不足となる演算部分を FPGA にオフロードする。また CPU よりも効率的な GPU・FPGA 間データ移動を実現できれば、GPU+FPGA 演算加速型ビッグデータ処理をさらに加速させることが可能である。ただ、ビッグデータ分析ワークロードに対する検討は十分に行なわれていない。

そこで本研究では GPU および FPGA を活用したビッグデータ分析の高速化に取り組む：1) GPU/FPGA もしくは両者を組み合わせることによって性能の改善が見込まれるビッグデータ分析ワークロードの検討、2) GPU および FPGA を連携したビッグデータ分析処理高速化手法の検討、3) プロトタイプシステムの実装および性能評価

### 2. 研究成果の内容

今年度の成果の概要は以下の通り。

#### ■ FPGA によるラベル付き有向グラフ上の正規パス問合せの高速化

正規パス問合せ (regular path query) とは、ラベル付き有向グラフから、問合せとして与えられた正規パスを経由して到達可能な全ての頂点を列挙する問合せである。近年のデータ分析での対象データの大規模化の傾向から、大規模グラフに対する RPQ の高速化が期待されている。そこで本研究では FPGA を利用した RPQ 処理のアクセラレータを開発した (図 1)。

FPGA 上に比決定性有限オートマトン (NFA) 評価器を構築し、複数の探索を並列に実行することで高速化を実現した。性能調査のための実験では、RPQ で指定されるパスの長さに比例した実行時間

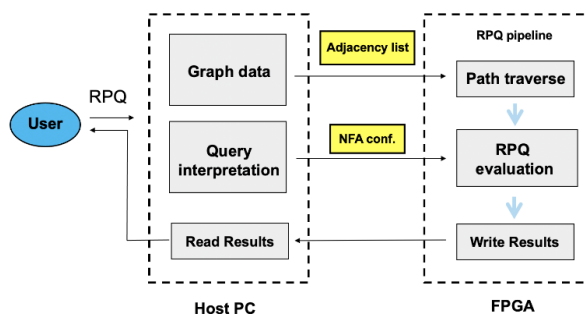


図 1 FPGA による RPQ アクセラレータ

を示し、比較手法と比べ最大で約 3 桁の高速化が得られた。また、あるグラフに対して同一のパス長の RPQ を実行する場合には実行時間がほぼ一定になることや和演算や繰り返し演算などを含む様な複雑な RPQ に対しても実行時間が影響されないといった特徴を確認した。

### ■ GPU による 5 ノードサブグラフ数え上げの高速化

サブグラフ数え上げは、対象となるグラフに対して、特定の（数ノードからなる）グラフ（パターンと呼ばれる）がサブグラフとして出現する回数を求める問題である。本研究では、パターンのサイズに上限を設定し、サイズが上限となるあらゆるパターンに対して数え上げを行なう問題に取り組んだ。このサブグラフ数え上げは、グラフ分析の基本的な手法の一つであり、グラフの特徴を計算する手法の一つとして、バイオ情報学等の分野で利用されている。これまで厳密解を求める実用的なアルゴリズムは 4 ノードまでが上限であったが、Pinar らによる ESCAPE が、5 ノードサブグラフ数え上げに対してはじめて実用的なアルゴリズムを提案した。しかしながら GPU による処理の並列化は考慮されていなかった。

そこで本研究では、この問題に対して ESCAPE をベースとして、GPU による並列化を適用し高速化を行なった。具体的には、ESCAPE が提案しているグラフの有向グラフへの変換およびパターン分割による数え上げのアプローチを踏襲しつつ、並列化が容易な分割パターンの数え上げの部分 GPU 上で並列に実行する手法を開発した。特に、GPU 上でのメモリ効率の良いデータ構造を利用したグラフの表現（図 2 (a) および (b)）を適用するとともに、必要に応じてエッジを逆に辿るためのデータを保持したり（図 2 (c)）、分割

パターンのためのデータ構造を導入（図 2 (d)）することで効率の良い処理を実現した。その結果、ESCAPE に対して 5 ノードサブグラフ数え上げにおいて 4 倍から 10 倍の高速化を実現した。

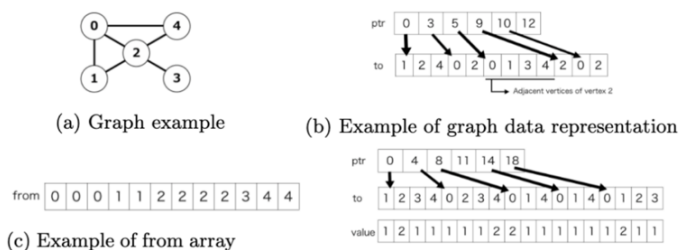


図 2 GPU による 5 ノードサブグラフ数え上

### 3. 学際共同利用が果たした役割と意義

最新鋭の GPU および FPGA を混載した計算機環境は構築や整備・運用が高コストである。そのような環境を提供していただいたことではじめて、本研究を推進することが可能となった。

### 4. 今後の展望

今回は単一 FPGA の利用にとどまったが、今後は、RPQ 処理についてマルチ FPGA 化に取り組みたい。さらに、GPU および FPGA を最適に組合せたビッグデータ処理についても検討を進めていく予定である。

5. 成果発表

(1) 学術論文

1. Shuya Suganami, Toshiyuki Amagasa, Hiroyuki Kitagawa, "Accelerating all 5-vertex subgraphs counting using GPUs", Proc. 31st International Conference on Database and Expert Systems Applications (DEXA2020), pp. 55-70, September 14-17, 2020.

(2) 学会発表

1. 三浦賢人, 天笠俊之, 北川博之, "FPGA を用いたラベル付きグラフに対する正規パス問合せアクセラレータの改良", 第 13 回データ工学と情報マネジメントに関するフォーラム (DEIM 2021), B13-3, 2021 年 3 月 1 日～3 月 3 日.
2. 菅波柊也, 天笠俊之, 塩川浩昭, 北川博之, "GPU 並列 5 ノードサブグラフカウンティング手法の改良", 第 13 回データ工学と情報マネジメントに関するフォーラム (DEIM 2021), B13-4, 2021 年 3 月 1 日～3 月 3 日.

(3) その他

使用計算機	使用計算機 に○	配分リソース※	
		当初配分	追加配分
Cygnus	○	10000	-
Oakforest-PACS			
※配分リソースについてはノード時間積をご記入ください。			