

氏名	小澤 佑介		
学位の種類	博士(工学)		
学位記番号	博甲第 7933 号		
学位授与年月日	平成 28年 9月 23日		
学位授与の要件	学位規則第4条第1項該当		
審査研究科	システム情報工学研究科		
学位論文題目	A Study on GPU-Accelerated Data Analysis Techniques (GPUによるデータ分析処理の高速化に関する研究)		
主査	筑波大学 教授	理学博士	北川 博之
副査	筑波大学 教授	博士(理学)	高橋 大介
副査	筑波大学 教授	学術博士	和田 耕一
副査	筑波大学 准教授	博士(工学)	天笠 俊之
副査	東京工業大学 教授	博士(情報科学)	宮崎 純

論文の要旨

大規模データの増加とともに、それらに対する高速なデータ分析に対する要求が高まっている。一方で、グラフィックス処理のための専用プロセッサである GPU を汎用的な処理に用いる GPGPU (general-purpose computing on graphics processing units) が注目されており、さまざまな処理の高速化に用いられている。GPU は多数のスレッドを利用した高い並列処理により高い処理性能が期待されるものの、通常の CPU とは異なったアーキテクチャを持つことから、その活用には以下のような技術的課題が存在する。すなわち、スレッド、スレッドブロック等の階層構造を意識した並列化、グローバルメモリ、ローカルメモリなどのサイズやレイテンシの異なる複数種類のメモリの利用、スレッド間の負荷均衡化である。

これらの問題に対し、本学位論文では、データ分析において重要な4つの処理を対象に GPU による高速化の手法を提案している。対象としている処理は、不確実データに対する頻出アイテム集合マイニング、比較ソート、キャノピークラスタリング、グラフクラスタリングである。提案手法は、1) データアクセスを効率的に行うためのデータ構造、2) 並列データ処理プリミティブの活用、3) 性能改善の上でキーとなるアルゴリズム上のコンポーネントの抽出、という3つの共通的なアプローチに基づくことで、4つの処理の高速化を実現している。

不確実データに対する頻出アイテム集合マイニングでは、異なるアイテム集合ごとにその出現確率を計算する必要がある。このため、既存手法 pApriori アルゴリズムを元に、GPU 上での処理に適したデータ構造を提案するとともに、GPU の高い並列度とメモリ階層を生かした確率計算の高速化手法を提案した。その結果、マルチコア CPU での並列実装に比べて 5.5 倍の性能を達成した。

これまで GPU を利用したソートアルゴリズムは多数提案されているが、スレッド間の負荷分散や大規模データに対する性能劣化などの問題があった。本研究ではサンプルソートと併合ソートを組み合わせた新たなアルゴリズムを提案している。サンプルソートの手法をベースに負荷分散を考慮したデータ分割を行い、分割済みデータの処理を併合ソートに基づいて行う。これにより、GPU による既存アルゴリズムに対して 32%の性能改善を達成した。

キャノピークラスタリングは、大規模データに対するクラスタリングの前処理として位置づけられるが、GPU を利用した処理はこれまでほとんど研究されてこなかった。本研究では、キャノピークラスタリングに適したデータ構造やデータ並列プリミティブを活用したアルゴリズムを提案した。さらに空間グリッドを利用した距離計算の削減についても提案した。その結果、8 コア CPU での並列実装に比べて 2.5 倍の性能を達成した。

グラフクラスタリングでは、大規模なグラフに対しても精度の良いクラスタリングが可能とされるラベル伝搬法を対象に、GPU の並列性能を生かしたアルゴリズムを提案している。特に、実世界の多くのグラフでは、頂点ごとに次数が大きく異なる。このため、単純な頂点分割ではスレッド毎の負荷が偏ってしまい、性能が改善しないという問題があった。これに対応するため、スレッドごとの負荷を均衡化するようなアルゴリズム上の工夫を行い、CPU 実装に比べて平均 30 倍の高速化を達成した。

審 査 の 要 旨

【批評】

ビッグデータが一般化し、情報システムには、膨大なデータに対して複雑な分析処理を高速に実行する能力が求められている。このような状況において、GPU は大規模データに対して高速な分析処理を可能にする強力な手段として注目されている。一方、その性能を引き出すには、GPU のアーキテクチャを考慮したデータ構造や並列アルゴリズムの設計および実装が必要である。本学位論文では、不確実データに対する頻出アイテム集合マイニング、比較ソート、キャノピークラスタリング、グラフクラスタリングを対象に、GPU を用いた高速化のための共通的なアプローチと、それに基づく具体的手法を提案している。さらに提案手法の有効性を実験的に示しており、CPU や従来手法に比べて大幅な高速化を達成している。これらの点において、情報工学上の顕著な貢献が認められる。今後は、より大規模なデータへの対応や、CPU やその他の異種プロセッサが混在した環境での高速化などへの発展が期待される。

【最終試験の結果】

平成 28 年 8 月 1 日、システム情報工学研究科において、学位論文審査委員の全員出席のもと、著者に論文について説明を求め、関連事項につき質疑応答を行った。その結果、学位論文審査委員全員によって、合格と判定された。

【結論】

上記の学位論文審査ならびに最終試験の結果に基づき、著者は博士（工学）の学位を受けるに十分な資格を有するものと認める。