

論文 / 著書情報
Article / Book Information

題目(和文)	メニーコアアーキテクチャにおける疎行列計算の性能最適化
Title(English)	Performance Optimization of Sparse Matrix Kernels for Many-core Architectures
著者(和文)	長坂 侑亮
Author(English)	Yusuke Nagasaka
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第11066号, 授与年月日:2019年3月26日, 学位の種別:課程博士, 審査員:松岡 聡,遠藤 敏夫,増原 英彦,額田 彰,横田 理央
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第11066号, Conferred date:2019/3/26, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	審査の要旨
Type(English)	Exam Summary

(博士課程)

論文審査の要旨及び審査員

報告番号	甲第	号	学位申請者氏名	長坂 侑亮		
論文審査 審査員		氏名	職名		氏名	職名
	主査	松岡 聡	特任教授		横田 理央	准教授
	審査員	遠藤 敏夫	教授	審査員		
		増原 英彦	教授			
		額田 彰	准教授			

論文審査の要旨 (2000 字程度)

本論文は「Performance Optimization of Sparse Matrix Kernels for Many-core Architectures (メニーコアアーキテクチャにおける疎行列計算の性能最適化)」と題し、数値計算やグラフ解析、機械学習などの主要な分野において処理性能の向上が求められる疎行列関連の計算を対象に、近年飛躍的に性能が向上している GPU 等のメニーコアプロセッサを用いて高速化するための最適化手法を提案するもので、英文で全 8 章から構成されている。

第 1 章「Introduction」では本論文の導入として、本研究が対象とするメニーコアアーキテクチャにおける疎行列計算の性能上の問題点を述べ、それらを解決するための最適化手法の提案などの貢献を端的に示し、最後に本論文の構成を述べている。

第 2 章「Background」では本研究の背景として、メニーコアプロセッサ及び疎行列計算と基礎的な疎行列格納方法等を述べている。

第 3 章「Improving Cache Locality in Sparse Matrix Vector Multiplication」では、疎行列ベクトル積計算におけるキャッシュ再利用性の問題を性能分析によって明らかにし、提案手法である NUS フォーマットを説明している。NUS フォーマットは疎行列ベクトル積計算での入力ベクトルへのメモリアクセスの局所性向上を可能とする。代表的なメニーコアプロセッサである NVIDIA GPU と Intel Xeon Phi にて性能評価を行い、その有効性を確認している。

第 4 章「Bandwidth Reducing Algorithm for Sparse Matrix Vector Multiplication」では、疎行列ベクトル積計算におけるメモリアクセス量の削減を目的とした疎行列の圧縮手法である AMB フォーマットを提案している。NUS フォーマットによって実現された入力ベクトルへのメモリアクセスの局所性向上を維持した上で、メモリアクセス量増加の要因となるインデックスデータの効果的な圧縮を実現している。GPU にて詳細な性能評価を行った結果、最大 1.4 倍の性能向上を達成したことを確認している。

第 5 章「Optimization Strategy of Sparse General Matrix Matrix Multiplication on GPU」では、GPU における疎行列同士の積計算を対象とした性能最適化手法を提案している。処理量に基づく GPU 資源の割当によって負荷不均衡の改善を達成し、適切なサイズのハッシュテーブルの構築によってメモリアクセスの局所性改善を実現している。GPU にて詳細な性能評価を行った結果、最大 4.4 倍の性能向上を達成し、メモリ使用量の削減によって巨大な入力に対しても処理が可能であることを確認している。

第 6 章「Optimization and Performance Analysis of SpGEMM on Intel Architectures」では、Intel Xeon Phi 等における疎行列同士の積計算の性能の要因をマイクロベンチマークによって明らかにし、疎行列同士の積計算の性能最適化手法と性能モデルを提案している。詳細な性能評価によって、提案手法の有効性と性能モデルの妥当性を確認しており、入力データに依存する諸手法の優位性について考察を行っている。また、グラフ解析を行うアプリケーションでの評価を行っており、提案手法の優位性を示している。

第 7 章「Batched Sparse Matrix Multiplication」では、小疎行列を対象とした疎行列積計算の最適

化手法を提案している。各疎行列積計算の高速化手法である Sub-Warp-Assigned SpMM と、複数の疎行列積計算を纏めることにより高い実行効率を実現する Batched SpMM を提案している。機械学習領域における Graph Convolution を行うアプリケーションにて、Batched SpMM を導入した場合の評価を行っており、最大 1.64 倍の性能向上を確認している。

第 8 章「Conclusion」では、本論文の総括を述べている。

以上のように、本研究はメニーコアプロセッサで高性能な疎行列計算を実現するための手法について提案し、また実験によってその有効性を確認しており、理学上貢献するところ大である。よって、本論文は博士（理学）の学位論文として十分価値があるものと認める。

注意：「論文審査の要旨及び審査員」は、東工大リサーチポジトリ(T2R2)にてインターネット公表されますので、公表可能な範囲の内容で作成してください。