

論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Cross-Architecture Performance Prediction Using Collaborative Filtering
著者(和文)	Salaria Shweta
Author(English)	Shweta Salaria
出典(和文)	学位:博士(学術), 学位授与機関:東京工業大学, 報告番号:甲第11318号, 授与年月日:2019年9月20日, 学位の種別:課程博士, 審査員:松岡 聡,遠藤 敏夫,脇田 建,額田 彰,横田 理央
Citation(English)	Degree:Doctor (Academic), Conferring organization: Tokyo Institute of Technology, Report number:甲第11318号, Conferred date:2019/9/20, 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	Shweta SALARIA	
論文審査 審査員		氏 名	職 名		氏 名	職 名
	主査	松岡 聡	特任教授	審査員	横田 理央	准教授
	審査員	遠藤 敏夫	教授			
		脇田 建	准教授			
		額田 彰	准教授			

論文審査の要旨 (2000 字程度)

本論文は「Cross-Architecture Performance Prediction Using Collaborative Filtering (異なるアーキテクチャにおける協調フィルタリングを用いた性能予測)」と題し、マルチコア CPU および GPU の異なるアーキテクチャ間の性能モデルの構築に関して、協調フィルタリングの手法が実用的かつ効果的であることを提案し、示すものである。この要旨は本論文の構成を要約して述べたもので、英文で全 6 章から構成されている。

第 1 章「Introduction」では、本論文の背後にある動機を示し、異なるアーキテクチャを持つシステムで予測モデルを構築するための主な課題を紹介し、それらの課題に対処するための本論文の貢献を列挙している。

第 2 章「Background」では、本研究の背景として、情報推薦システム (information recommendation system) で使用される様々な種類の協調フィルタリングの手法について説明している。また、予測モデルの重要な要素である 2 つの概念、潜在特徴 (latent feature) と埋め込み (embedding) について述べている。

第 3 章「Performance Evaluation」では、スーパーコンピュータとクラウドプラットフォームでアプリケーションを実行する際の性能のトレードオフを検証するために、3 つの異なるシステムを用いて 2 つの科学系ベンチマークの性能を最適化して評価している。高性能計算に特化した AWS (Amazon Web Service) C4 インスタンスの計算性能を、本学の TSUBAME スーパーコンピュータと比較し、計測した計算や通信性能やその変動性、およびワークロードに最適なシステムを選択するためには性能モデルの構築に焦点を当てる必要性があることを説明している。

第 4 章「Predicting CPU Performance」では、異なる CPU の性能を予測するために協調フィルタリングの手法を使用することを提案し、その実用性を検証している。まず、異なるプロセッサアーキテクチャを評価するための関連研究を紹介している。次に、性能データセットを構築し、マルチコアシステム上の異なるプロセッサ間の性能を予測するために 2 つの協調フィルタリングモデルを提案し、実装している。特に、アプリケーションとシステム間の相互作用を捉える潜在的機能をモデル化するた

めに行列分解に基づく協調フィルタリング手法が有効であることを示している。最後に、主成分分析と階層型クラスタリングを用いて統計分析を実行することで、本研究に使用したモデルが重要な特徴量を学習できることを検証し、これらの手法がシステムとベンチマークの相関をモデル化するのに相応しいことを示している。

第5章「Predicting GPU Performance」では、GPUの性能をモデル化するために、第4章で提案された行列分解に基づく協調フィルタリング手法を拡張している。9つの異なるアプリケーション領域とNVIDIA社の4つの異なる世代（Kepler、MaxWell、Pascal、Volta）にわたる7種類のGPUからのワークロードで構成されるデータベースを構築することによって、モデルを経験的に評価している。また、様々なGPUを評価する関連研究についても議論している。次に、行列分解の手法の限界を論じ、予測精度をより向上させるためにハードウェアとプログラムの複雑な相互作用を学習することができる、多層パーセプトロンを使用した新しいモデルを提案している。最後に、多層パーセプトロンを使用した埋め込みの学習が性能予測に関し、行列分解法と比較してより正確で優れていることを示している。

第6章「Summary」では、本研究の主な貢献を要約し、将来の方向性についていくつかの洞察を与えている。

以上のように、本研究は機械学習を使用して異なるプロセッサおよびHPCアクセラレータ上の性能を予測する複数の方法を提案および評価することで、新しいHPCシステム的设计に貢献し、それらが設計される以前にシステムの評価を単純化することを可能にし、また実験によってその有効性を確認しており、学術上貢献するところ大である。よって、本論文は博士（学術）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。