

氏 名 劉 雨

学位(専攻分野) 博士(情報学)

学位記番号 総研大甲第 1721 号

学位授与の日付 平成26年9月29日

学位授与の要件 複合科学研究科 情報学専攻
学位規則第6条第1項該当

学位論文題目 A Computational High-Level Parallel Programming Framework
for Data-Intensive Computing

論文審査委員 主 査 教授 胡 振江
教授 中島 震
准教授 吉岡 信和
助教 日高 宗一郎
教授 岩崎 英哉 電気通信大学

論文内容の要旨
Summary of thesis contents

Recent years parallel programming for efficiently processing large-scale data becomes a crucial problem in practice. However, writing parallel programs has proven to be a difficult task over the years, requiring various specialized knowledge and skills. Traditional parallel programming models and frameworks like MPI, OpenMP or Pthreads, do support writing efficient parallel programs, but for data-intensive processing, programming under these frameworks usually leads to overly complex, non-portable, and often unscalable and unreliable code.

This thesis reports a novel calculational approach to constructing a high-level parallel programming framework for data-intensive computing. The main contribution of this thesis is a high-level parallel programming framework for data-intensive computing. This framework makes it more systematic to generate efficient parallel programs for processing large scale data. It has a set of libraries that not only support declarative descriptions of data-intensive computing over various data structures such as lists, sets, trees and graphs, but also automatically transform sequential programs into efficient and scalable MapReduce programs. Moreover, this programming framework enforces deterministic semantics that simplify the tasks of composing, optimizing, porting, reasoning about, debugging, and testing parallel programs.

The thesis consists of seven chapters. Chapter 1 is the introduction of the thesis, explaining the research background, objective, and main contributions of this thesis. The other Chapters of this thesis are presented in three parts.

In the first part (Chapters 2 and 3), an approach to resolving problems of the data model list is introduced. In particular, Chapter 2 introduces algebras for various data structures and homomorphisms on the algebras. Our high-level parallel programming principle is based on the structures of data and algebras for such data structures. There are deep relationships between the algebraic data structures and the computations (i.e., the problem specifications as well as the algorithms) defined on these data structures. We make use of the mathematical properties (e.g., associativity, commutativity, distributivity and fusibility of the operators inside the definitions of algebras) of these algebras to construct, transform and optimize programs. In Chapter 3, we introduce our calculational approach to resolving programming problems that are based on the data model of lists or sets.

The second part (Chapterd 4 and 5) of the thesis describes our contributions on resolving programming problems for certain combinatorial optimization problems on graphs. In Chapter 4, we introduce our approach to parallel programming on large scale data that can be represented as special trees called tree decompositions of graphs, by which we can represent a large graph as a tree and applying efficient dynamic programming algorithms on such tree to resolve many NP-hard optimization problems.

(別紙様式 2)
(Separate Form 2)

Tree decompositions act as a bridge that brings the unstructured graph data into our framework. In Chapter 5, an automatic program transformation approach to some optimization problems on large-scale graphs is introduced.

Implementations, evaluations and conclusions are covered in the last part (Chapters 6 and 7). In Chapter 6, we explain the implementation of the libraries inside our framework. Finally, a summary of this thesis and an outline of future work are given in Chapter 7.

(別紙様式 3)
(Separate Form 3)

博士論文の審査結果の要旨
Summary of the results of the doctoral thesis screening

審査委員全員出席の下、博士学位請求論文(以下、「論文」)の内容についてスライドを用いて約 45 分間の口頭発表を行った後、質疑応答を 30 分間行った。

論文では、データ集約的計算とそのためのプログラミングモデルを研究背景に、MapReduce とその上に構築された既存の並列プログラミングモデルの問題点を解決するための、プログラム演算 (program calculation) による新しいプログラミングモデルの提案・実現・評価について論じている。新しく提案された高水準プログラミングモデルの特徴は、ユーザがデータ集約的計算を宣言的に記述できるプログラミングインターフェースを提供するとともに、プログラム演算により宣言的な記述から効率のよい MapReduce 計算に自動的に変換できるという点である。また、リストや集合といった平坦なデータだけではなく、木やグラフといった複雑な関係をもつデータも統一的な枠組みで議論できることを示している。

論文は英語で記述されており、全 7 章から構成されている。

第 1 章はイントロダクションである。研究の背景、研究目的、主要な貢献など論文全体の構成を述べている。

第 2 章では、リストを含む様々な代数型データとその上の準同型を紹介し、準同型とその合成によって様々なアルゴリズムを記述することが可能となり、さらに準同型上の代数性質を利用しプログラム演算で最適化できることを示している。

第 3 章では、リストに特化して、リストを処理する計算問題を解決するための記述言語、宣言的記述から並列スケルトンに変換するルール、及び MapReduce によるスケルトンの実現方法を示している。

第 4 章と第 5 章では、リスト上の手法を拡張し、グラフ上の組合せ最適化問題を解く並列化手法を論じている。まず第 4 章では、有界な木幅を持つグラフについては、グラフを木構造に分解し、また木構造をリストで表現することにより、グラフ上の問題をリスト上の問題に帰着する方法を提案している。これにより、多くのグラフアルゴリズムを動的計画法に基づいて効率的に MapReduce で実装でき、大規模グラフの取り扱いが可能となる。次に第 5 章で、有界な木幅を持たないグラフを有界な木幅を持つグラフで近似する方法を提案し、それにより任意のグラフを処理することが可能になることを論じている。また、その提案を実際の Target Set Selection 問題へ応用し、大規模データを用いた提案手法の有効性を示している。

第 6 章では、新しい並列プログラミングフレームワークの実装と評価について述べている。特に、リスト上の準同型計算、蓄積計算、Generate-Test-Aggregate 計算パターンの MapReduce 実装方法を示すとともに、実験によってその有用性を示している。

第 7 章は論文のまとめと今後の課題である。データ集約的計算を対象に、提案した高水準並列プログラミングフレームワークが新規性があり有効であると結論づけるとともに、将来の課題を論じている。

(別紙様式 3)

(Separate Form 3)

なお、この論文成果は国内外学会論文誌 3 件、国際会議論文 3 件として公開されている。
また、開発したシステムもウェブで自由にダウンロードできるようになっている。

審査会において、出願者は上述の内容に沿って説明を行い、そのあと審査委員との質疑応答を行った。質疑応答では、論文及び口頭発表の内容に関して、ユーザのプログラミングインターフェース、グラフ処理とリスト処理との関係を中心に質問があり、的確な回答がなされた。

以上に基づき審査した結果、5 名の審査委員全員一致で、本学位請求論文は学位を授与するのに十分なレベルであるものと判定された。