

審査の結果の要旨

論文提出者氏名 イコ プラムディオノ

本論文は「Parallel Platform for Large Scale Web Usage Mining(大規模ウェブログマイニングの為の並列プラットフォームに関する研究)」と題し、Web ユーザの行動を記録した大容量ウェブアクセスログを解析するための並列ウェブマイニングプラットフォームの構築法を明らかにすることを目的とし、ウェブログマイニングの効率の良い並列アルゴリズムを提案するとともに、PC クラスタ上に当該アルゴリズムを実装、性能評価することによりその有効性を明らかにしており、8章から構成される。

第1章は”Introduction”(序論)であり、本研究の背景、目的、及び位置付けについて述べている。

第2章は”Web Usage Mining”(ウェブログマイニング)と題し、ウェブマイニングにおける既存の関連研究、アクセスログのマイニング処理全体の概要、および頻出パターンマイニング、時系列パターンマイニング等の代表的なマイニング技術を説明している。

第3章は“PC Cluster Based Web Usage Mining Platform”(PC クラスタ上のウェブログマイニングプラットフォーム)と題し、大規模アクセスログの柔軟な処理を可能とするための高性能プラットフォームに対する要件を整理し、当該要件を満たすPC クラスタによるウェブマイニングシステムの構成を提案している。

第4章は“Parallel Algorithm to Mine Frequent Patterns”(頻出パターンマイニングの並列アルゴリズム)と題し、頻出パターンマイニングアルゴリズム FP-growth の並列化手法を提案している。FP-growth は、トランザクションデータベースを FP-tree と呼ばれる主記憶上データ構造に圧縮し展開することにより、広く利用されてきた Apriori アルゴリズムに比べ、大幅な高速化を達成しているが、木状の複雑なデータ構造を操作するため、PC クラスタ等の無共有型並列計算機上での並列化は困難と考えられ従来試みられて来なかった。本章では、FP 木に基づく新しい並列アルゴリズムを提案している。粒度を実行時に調整可能とし、ノード間の動的負荷分散機構を提案している。加えて、FP-tree 枝の重複を減らし、実行時メモリ消費量を最適化する手法も提案している。

第5章は“Parallel Algorithm to Mine Web Access Sequence Patterns”(並列ウェブアクセスシーケンスパターンマイニングの並列アルゴリズム)と題し、時系列パターンの一種であるウェブアクセスパターン(WAP)の並列マイニング手法を提案している。第4章で提案された手法を拡張し、WAP-tree のノードにあるアイテムのカウント情報を直接操作する新しい高速なシーケンスマイニングアルゴリズムを提案している。

第6章は“Performance Evaluation of Parallel Algorithms on PC Cluster”(PC クラスタ上の並列アルゴリズム性能評価)と題し、第4章と第5章で提案された並列ウェブマイニング方式をPC クラスタ上で実装し、性能評価の結果を詳細な実行トレースを解析することにより、提案手法の有効性を定量的に明らかにしている。頻出パターンマイニングの場合、アプリアリを基本とする従来の並列アルゴリズム HPA に比べ一桁以上高速化されており、32台の実行ノードで23倍の高速化が達成出来ることを示している。又、動的に FP-tree を再構築しメモリ消費量を最適化することにより、16ノード以上の場合において、30%以上のメモリ空間の節約が可能であることが示されている。

第7章は“Applying the Platform on Logs of Location Aware Portal Sites”(位置情報サービスポータルサイトのログへの適用)と題し、提案されたプラットフォームを位置情報ポータルサイト Mobile

Info Search(MIS)及び電話番号情報サイト iTOWNPAGE のアクセスログに適用することにより推薦システムなどを構築し、提案したウェブマイニングシステムの有効性を明らかにしている。

第8章 “Conclusion”（結論）は、本論文の成果と今後の課題について検討している。

以上これを要するに、本論文は、大規模ウェブログマイニングを可能とする PC クラスタを用いたプラットフォームの構築を目的とし、FP 木に基づく新しい並列頻出パターンマイニングアルゴリズムを提案するとともに、実装によりその高い性能を確認し、実際の大容量アクセスログを用いることにより提案プラットフォームの有効性を明らかにしており、電子情報工学上貢献するところが少なくない。

よって本論文は博士（工学）の学位請求論文として合格と認められる。

「審査の結果の要旨」の概要

1. 課程・論文博士の別 課程博士
2. 申請者氏名（ふりがな） イコ プラムディオノ（いこ ぷらむでいおの）
3. 学位の種類 博士（工学）
4. 学位記番号 博情 第 号
5. 学位記授与年月日 平成 年 月 日
6. 論文題目 Parallel Platform for Large Scale Web Usage Mining
(和訳 大規模ウェブログマイニングの為の
並列プラットフォームに関する研究)
7. 審査委員会委員 (主査) 東京大学 教授 喜連川 優
教授 田中 英彦
教授 安達 淳
教授 相田 仁
教授 坂井 修一

8. 提出ファイルの仕様等

	ファイル名	使用アプリケーション	OS
使用文書ファイル名	イコプラムディオノ.doc	word2000	windowsXP
テキストファイル名	イコプラムディオノ.txt		

画像ファイル

(ある場合のみ)