

自己評価報告書

平成 23 年 4 月 12 日現在

機関番号：10101
研究種目：基盤研究(B)
研究期間：2008～2011
課題番号：20300051
研究課題名(和文) 超大規模な単一実メモリ空間を活用するデータベース解析処理アルゴリズムの研究
研究課題名(英文) Research on database analysis algorithms using very large-scale monolithic memory space
研究代表者
湊 真一(MINATO SHIN-ICHI)
北海道大学・大学院情報科学研究科・教授
研究者番号：10374612

研究分野：総合領域
科研費の分科・細目：情報学・知能情報学
キーワード：知識発見とデータマイニング、二分決定グラフ

1. 研究計画の概要

データベース分野では、当初はハードディスク内にデータを格納することが一般的だったが、2000年以降、中規模クラスの実用的なデータベースが計算機のメモリに納まるようになってきた。そこで本研究課題では、超大規模な単一実メモリ空間を持つ計算機に、二分決定グラフ(BDD)と呼ばれる巨大な圧縮データ構造を構築して、全てを高速アクセス可能なメモリ上に格納し、高速にデータベース解析処理を行う手法を提案し、その研究実用化を進める。

2. 研究の進捗状況

以下の3項目について研究を進めた。

(1) BDD/ZBDDに基づくデータベース解析処理アルゴリズムの効率化と一般化の研究

組合せアイテム集合の出現回数(頻度)をZBDDによりインデックス化してメモリ上に生成・保持する効率的なアルゴリズムを既存手法と比較し、その効果を明らかにした。頻出アイテム集合を生成することを知識概念の抽出・学習過程と捉えて、BDDを用いてコルモゴロフ複雑度を求めた場合の処理効率や誤差を、理論と実装の両面から調べ、機械学習や自動クラスタリングへの応用可能性を調べ、BDD/ZBDDの知識表現としての学術的側面を追究した。さらに、毎日のデータベースから抽出した日毎の頻出パターンをZDDの系列として表現し、ZDD動詞の演算処理により特徴的な変化を効率よく発見する手法を提案した。「共起成分の含意関係」(Minato, 2007)と呼ぶ新しい含意関係をさらに発展させ、誤差を許した共起含意関係を抽出するアルゴリズムの研究を進めた。

(2) 超大規模な単一実メモリ空間を活用するアルゴリズム実装技術の研究

これまでに開発した各種のBDD/ZBDD処理アルゴリズムを上記のメモリサーバPCに実装し、性能評価と課題の抽出を行った。大規模なメモリ空間を使い切るBDD/ZBDD処理パッケージの保守管理を継続的に行い、利用ノウハウを蓄積するとともに、プログラムやドキュメント、評価用ベンチマーク例題などを国内外に随時公開した。2010年度には、当初計画していた通り、256GBの単一実メモリを持つ汎用PCを3台導入することができたので、今後、このPCのメモリを使い切る基盤ソフトウェアの開発を目指す。

(3) 本基盤技術の実データへの応用と性能評価、および課題のフィードバック

現実のWeb空間に存在するデータを題材に、本基盤技術の有効性を評価する。Web上の文献情報の検索と類似文献のクラスタリング、XMLのような木構造を持つデータの解析、さらに長大なテキスト情報の検索と知識発見を検討した。生命情報科学分野におけるウイルスの変異予測や遺伝情報からのデータマイニングへの応用法の検討を進めた。

3. 現在までの達成度

②おおむね順調に進展している。

(理由)

初年度より、国際会議発表等の研究成果が順調に出ている。また当初の計画通り、2010年度に世界最大級の単一実メモリ空間を備えた汎用PCを導入することができ、最終年度に向けた研究環境が整った。

4. 今後の研究の推進方策

2010年度に導入した世界最大級の単一実メモリ空間を備えた汎用PCを用いて、実験と

評価を行い、4年間の研究成果をまとめる予定である。

5. 代表的な研究成果

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計8件)

- (1) 湊 真一: “BDD/ZDD を基盤とする離散構造と演算処理系の最近の展開,” 電子情報通信学会 基礎・境界ソサイエティ Fundamentals Review, Vol. 4, No. 3, pp. 224~230, 2011. (招待, 査読有)
- (2) Shin-ichi Minato and Takeaki Uno: “Frequentness-Transition Queries for Distinctive Pattern Mining from Time-Segmented Databases,” Proc. of 2010 SIAM International Conference on Data Mining (SDM2010), pp. 339~349, 2010. (査読有)
- (3) 湊真一, 有村博紀: “「おめでとうソサイエティ論文賞」ゼロサプレス型二分決定グラフを用いたトランザクションデータベースの効率的解析手法,” 電子情報通信学会 情報・システムソサイエティ誌, Vol. 15, No. 3, p. 15, 2010. (招待, 査読無)
- (4) Shin-ichi Minato, Takeaki Uno, and Hiroki Arimura: “LCM over ZBDDs: Fast Generation of Very Large-Scale Frequent Itemsets Using a Compact Graph-Based Representation,” In Proc. of PAKDD 2008, LNAI 5012, Springer, pp. 234 - 246, 2008. (査読有)
- (5) Shigeru Yamashita, Shin-ichi Minato, and D. M. Miller: “DDMF: An Efficient Decision Diagram Structure for Design Verification of Quantum Circuits under a Practical Restriction,” IEICE Trans. Fundamentals, Vol. E91-A, No.12, pp. 3793-3802, 2008. (査読有)

[学会発表] (計13件)

- (1) Shin-ichi Minato: “Discrete Structure Manipulation for Discovery Science Problems,” 25th Annual International Symposium on Computer and Information Sciences (ISCIS 2010), 2010年9月22日, 英国・王立科学協会
- (2) Shin-ichi Minato: “Recent Topics on Decision Diagrams and Discrete Structure Manipulation” (招待講演), 9th International Workshop on Boolean Problems (IWSBP 2010), 2010年9月16日, ドイツ・Freiberg 大学

- (3) 岡崎佑太, 湊真一: “頻出パターン抽出アルゴリズム「LCM over ZDDs」の変数順序付けの影響に関する考察,” FIT-2009 IEICE/IPSJ 第8回 情報科学技術フォーラム, 2009年9月4日, 東北大学
- (4) Shin-ichi Minato: “Recent Topics on BDDs/ZDDs for Data Mining and Knowledge Discovery” (招待講演), The International Workshop on Data Mining and Statistical Science (DMSS 2009), 2009年7月7日, 京都大学
- (5) 齋藤高央, 湊真一, ツオイクマントーマス: “コルモゴロフ複雑性に基づく画像圧縮と分類に関する実験と考察,” FIT-2008 IEICE/IPSJ 第7回情報科学技術フォーラム, 2008年9月4日, 慶応義塾大学湘南藤沢キャンパス

[図書] (計2件)

- (1) Shin-ichi Minato: “Morgan & Claypool Publishers, “Data Mining Using Binary Decision Diagrams,” In T. Sasao and J. Butler, editor, “Recent Progress in Representation of Discrete Functions,” chapter 5, pp. 97-109, 2010.
- (2) S. Chawla, T. Washio, S. Minato, S. Tsumoto, T. Onoda, S. Yamada, A. Inokuchi, (Editors): Springer, “New Frontiers in Applied Data Mining,” PAKDD 2008 International Workshops, Osaka, Japan, May 20-23, 2008, Revised Selected Papers, Lecture Notes in Artificial Intelligence, Vol. 5433, 2009. (213 pages)

[その他]

アルゴリズムのバイブルとも呼ばれ世界中で読まれているKnuthの名著 “The Art of Computer Programming” (Vol.4, Fascicle 1, 2009年)において、湊が考案したデータ構造「ZDD」が項目として詳しく掲載された(日本人初)。