自己評価報告書

平成23年4月27日現在

機関番号：10101
研究種目：基盤研究（A）
研究期間：2008 ～ 2011
課題番号：20240014
研究課題名（和文）大規模知識基盤形成のための次世代半構造マイニング技術の研究

研究課題名（英文）Next-Generation Semi-structured Data Mining for Large-Scale Knowledge Base Formation

研究代表者
有村 博紀（ARIMURA HIROKI）
北海道大学・大学院情報科学研究科・教授
研究番号番号：2022763

研究分野：情報科学

1. 研究計画の概要

本研究においては、ネットワーク上の大規模半構造データに在する知識をパターン化や規則としてとり出すことが可能な超高速半構造マイニングエンジンを開発し、これを現実の多様な半構造データに適用するための周辺技術を開発する。さらに、開発した基盤技術と周辺技術の実装を行い、インターネット上の大規模半構造データからの知識発見実験を行うことを目的とする。

2. 研究の進捗状況

（研究全体の進捗状況）平成20年度で、大規模知識基盤形成システムのための技術開発、基盤技術の開発、環境整備を行った。平成21年度で平成22年度は、調査整備に基づき、大規模知識基盤形成システムのための基盤および応用技術の開発と、その理論解析、最適化を行った。具体的には、これまでの3年間の研究期間で、次の項目に関して研究・開発を行った。

（1）超高速半構造マイニングエンジンの研究開発（有村・宇野・喜田）

高速な重み付けマイニングや、2次元画像、幾何グラフ、対応最適化データマイニング技術を開発した。超高速半構造マイニングエンジンの研究開発として、以上の要因に対する深さ優先方式の高速化を実現した。これに関して、連携研究者が情報処理学会平成22年度IPSJ論文賞井手賞奨励賞（Katoh, Hirata, Arimura, 2011.03.25）を受賞した。

（2）確率的情報処理スキームと半構造データマイニングの結合の研究（有村・喜田・湊・宇野）

ストリームハードウェア上の柔軟な高速半構造パターン照合方法（FPT2011）や、確率的な画像検索を用いたタグ照合問題（DS2008）を開発した。さらに、視覚データ分析上のテクニックモデルを導入し、動作レベルのアルゴリズムを開発した。実証実験として感染症の流行解析やジェンルタダリングの実証実験を行った（GIW2009）。連携研究者が情報処理学会平成22年度情報処理学会山下記念研究賞を受賞した（柳橋、2009年SIGBIO研究会）。

（3）半構造データマイニングの一般理論の構築（有村・宇野・湊）。従来の半構造データマイニング手法の構築を一般化し、問題が可能集合を基に図像データ構造を一般化する手法。データ・メタデータの統合ベースが求められている。

（4）大規模知識基盤形成システムのための半自動知識関係技術の研究開発（2009年・2010年）

IWOC2010), そのFPGAやGPUなどの並列ハードウェアへの実装方法 (FP2010) などの成果を得た（伊藤・喜田・湊・有村）（5）ZBDDに基づく高速な大規模知識索引技術を開発した（湊・喜田・有村）分担者の渋が開発したZBDDに基づく大規模知識索引技術に基づいて、データベースシステム系列から、知識索引と頻出アイテム集合成表を密結合したLCMoverZDD技術（Minato, Arimura, PAKDD2008）や、与えられた生起パターン表現をもと頻出アイテム集合成表手法の開発 (Minato, Uno, SDM2010)、大規模な系列データのための大規模知識索引技術SeqZDDの理論的解析とシステム構築等を行った。さらに分担者の渋は、巡回の離散構造上でのZDDの拡張に成功した (SAT2011, to appear)。一連の研究に関し、2010年信学会 情報・システムソサイエティ論文賞(湊,有村,2010.06.01)を受賞した。（6）研究開発と並行して、開発した知識発見技術のプロトタイプ実装を行った。さらに、理論解析と計算機実験による評価を行い、これらを元にさらなる最適化を行った。また、実装と計算機実験を行った（平成21年度）一連の研究開発と並行して、開発した知識発見技術のプロトタイプ実装を行った。さらに、理論解析と計算機実験による評価を行い、これらを元にさらなる最適化を行った。また、実装と計算機実験を行った（平成21年度）一連の研究開発と並行して、開発した知識発見技術のプロトタイプ実装を行った。さらに、理論解析と計算機実験による評価を行い、これらを元にさらなる最適化を行った。また、実装と計算機実験を行った（平成21年度）一連の研究開発と並行して、開発した知識発見技術のプロトタイプ実装を行った。さらに、理論解析と計算機実験による評価を行い、これらを元にさらなる最適化を行った。また、実装と計算機実験を行った（平成21年度）

3. 現在までの達成度

① 当初の計画以上に進展している。この4年間で、当初の予定通り、本研究の中核である超高速半構造マイニングアルゴリズムの開発と一般的設計手法の確立に成功しており、さらに、期待以上の結果として、生産性である時系列データでのグラフパターン発見や、知識索引との連携、大規模プロトタイプ実装等の成果が得られ、国際的な発表・出版を行うことができた。

4. 今後の研究の推進方策

研究項目(1)～(6)の各項目について、当初の計画以上に進展しているので、今後も研究計画を遂行する。特に、最終年度の2011年年度は、(1)を中心に、(2)では木構造上の確率モデルの効率よい学習、(3)では列挙可能性と計算効率の一般的特徴づけ、(4)では圧縮検索と超高速ビット並列検索、(5)では各半構造データへの知識索引の一般化、(6)プロトタイプの評価実験に重点を置いて研究開発を行う。

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おける下線)

〔雑誌論文〕（計 23 件）


〔学会発表〕（計 38 件）

(1) H. Arimura, Efficient Algorithms for Mining Frequent and Closed Patterns from Semi-structured Data, Proc. 12th Pacific-Asia Conf. on Knowledge Discovery and Data Mining (PAKDD2008), LNCS, Vol.5012, Springer, 2008.（招待講演）

[図書]（計 0 件）

[その他]