

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成24年5月15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・情報科学研究科・教授

研究者番号：20222763

研究成果の概要（和文）：本研究では、ネットワーク上の大規模半構造データからの知識獲得のための超高速な半構造マイニングエンジン技術と、これを現実の多様な半構造データに適用するためのさまざまな周辺技術を開発した。さらに、開発した技術の計算機上への実装を行い、大規模半構造データからの知識発見実験を行った。

研究成果の概要（英文）： In this research, we study efficient semi-structured data mining technologies and related knowledge federation and indexing technologies, which enables knowledge extraction from massive semi-structured data in the real world. Finally, we made a number of knowledge extraction experiments.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2008年度	14,800,000	4,440,000	19,240,000
2009年度	9,000,000	2,700,000	11,700,000
2010年度	9,000,000	2,700,000	11,700,000
2011年度	6,000,000	1,800,000	7,800,000
年度			
総計	38,800,000	11,640,000	50,440,000

研究分野：情報学

科研費の分科・細目：知能情報学

キーワード：知識発見とデータマイニング

## 1. 研究開始当初の背景

（1）1990年代のWorld Wide Web（ウェブ）に代表される高速ネットワークと情報処理技術の急速な発展によって、現在、科学技術や経済・社会活動に関する複雑かつ膨大な量の知識が、インターネット上に蓄積されつつある。例えば、インターネットのウェブ上には420万サイトを超越するウェブサイトが存在し、ウェブは人類が歴史上かつて体験したことがない膨大な知識の集積となっており、ウェブは巨大な知識ネットワークと

なりつつある。

（2）このウェブが持つ膨大な知識量のために、すでに生命科学分野やナノ技術分野などの知識集約的研究・産業分野では、一人の人間がこれらの膨大な知識を理解し、関連する多様な分野を横断的に研究を遂行することが困難な状況になりつつある。そのため、ウェブ上に集積された膨大な知識を関連付け、外在化して理解し、知識を取り出すための新しい情報アクセス技術の開発が緊急の課題となっている。

## 2. 研究の目的

データマイニングは、データベースに蓄積された一見無意味にみえる大量のデータから、知識として自明でない規則性やパターンを半自動的に取り出す方法についての科学的研究であり、1990年代初頭から顕在化し活発な研究が進んできた。しかし、その一方で、ウェブ空間の知識ネットワークは、(i)膨大な量の、(ii)多様な語彙と形式をもつ、(iii)非定型データの集積、すなわち大規模な半構造データであるので、従来のデータマイニング手法を、そのまま適用できない。

そのため、本研究では、大規模知識ネットワークを対象としたデータマイニング、すなわち知識ネットワークマイニングについて研究する。特に、実世界の多様な半構造データからの知識基盤形成を可能にするために、理論的な性能保障をもつ高速かつ頑健なマイニングアルゴリズムを開発する。そのために、鍵になる技術として、申請者等が開発してきた高速な最適化半構造データマイニングの枠組みを採用する(図1)。

本研究の特色として、(1)知識ネットワークマイニングの3つの基本的問題すべてを統一的に扱い、ウェブ上の大規模知識ネットワークマイニングを系統立てて研究し、(2)これをコンテンツベースの新しい大規模データアクセス手法として提案すること。(3)さらに、計算時間に徹底的にこだわり、アルゴリズム理論設計技法を駆使して、きわめて高速なアルゴリズムを追求する。

## 3. 研究の方法

申請者等は、これまでの研究で、木構造をもつ半構造データに対して、きわめて高速な最適化半構造データマイニング・エンジンの開発に成功している。そこで、本研究では、これまでに開発してきた半構造データマイニング・エンジンをより一般的なグラフ構造に拡張して、(A)知識ネットワークに対する最適化半構造データマイニング・エンジンを開発する。ここに、半構造マイニング・エンジンとは、与えられた木やグラフのような構造データの集合から、その共通の部分構造や特徴的な部分構造を高速に抽出する方法のことである(図1)。さらに、(B1)半自動的知識関係機構の研究および、(B2)大規模知識索引の研究という基本的な知識処理問題を定式化し、これらの3つの過程を統一的かつ有機的に結合し、半構造データマイニングを用いた知識ネットワークからの知識基盤形成システムの実現方式を明らかにする。

研究の進め方としては、実用と応用の両面から研究を進める。すなわち、各知識発見問題に対して効率よいアルゴリズムを開発する一方で、これらの問題を理論的に定式化し、

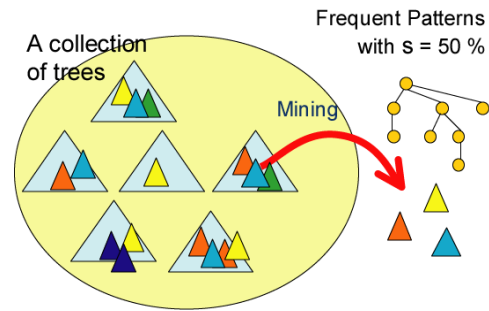


図1. 半構造データマイニング

その本質的な難しさを理論的に特徴付ける。とくに、申請者らが2000年初頭から研究してきたデータマイニングの理論的基盤と高速なアルゴリズムに基づいて、大規模半構造マイニング技術の開発を行う。

木構造を中心とする半構造マイニングは、「最右拡張法」に基づいて研究開発する。この最右拡張法は、2002年に申請者らが、米国や国内の研究グループ(Zaki, 2002; Nakano Inf. Proc. Letters, 2002)と独立に発見し、現在、多くの現代的グラフマイニング手法発展の基礎となった半構造マイニングの基盤技術である(Asai, Arimura et al., 2002)。これを、時系列上のグラフ構造や、幾何情報などのさまざまな半構造データへ拡張する(有村・宇野・平田)。

本研究で追求する極大半構造マイニングには、2003年に離散構造列挙アルゴリズムと伝統的マイニング手法の徹底した解析を元に開発された、理論的性能保障をもつ現在世界最速の代表元パターンアルゴリズムであるLCMの技術を基盤とする(宇野・有村, 2004年IEEE FIMIコンテスト総合優勝)。

さらに、LCM技術の半構造マイニングへの拡張については、2005年から推進した「半構造マイニング」に関する科研費特別推進研究(2005~2007, 代表:有村)で追求した、系列やグラフの自明でない部分族に対する世界初の理論的性能保障をもつ半構造マイニングアルゴリズムのアイデアを用いる(有村・宇野)。

半構造マイニングの重要な周辺技術である知識発見と知識処理の融合においては、分担者の湊が開発した大規模論理最適化技術であるゼロサプレスBDD(ZDD)を知識発見に導入することで、知識の圧縮と高速演算を援用した大規模知識索引技術を確立する(湊・有村・宇野)。このZDDは、計算機科学分野で最も権威がある教科書シリーズの一つであるD. E. Knuthの"The Art of Computer Programming"で日本人として初めて30ページにわたって大きく紹介されるなど、半構造マイニングの前処理と後処理の鍵となる最重要技術の一つである。

さらに高速圧縮照合技術(喜田)や、ネッ

トワーク上の情報連携技術(伊藤)を情報検索・知識連携との融合(喜田・伊藤)も行う。

#### 4. 研究成果

ネットワーク上の大規模半構造データに内在する知識をパターンや規則としてとりだすことが可能な超高速な半構造マイニングエンジン技術を開発し、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した。さらに、開発した基盤技術と周辺技術の実装を行い、実際の大規模半構造データ上の知識発見実験によってその性能評価を行った。

(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・平田)。

本研究項目では、高速な重み付きマイニングや、2次元画像、幾何グラフに対する半構造データマイニング技術を開発した。

具体的には、系列からのグラフ発見(エピソードマイニング)に関する研究では、2部グラフと多部グラフ、菱形エピソードに対する深さ優先方式の頻出エピソードマイニング手法の研究を行い(DS'09, PAKDD'09)、その成果の論文(TOD'10)は情報処理学会平成22年度船井若手奨励賞(Katoh他, 2011.03.25)を受賞した。

系列パターンに対する極大出現区間の概念を用いた高速な頻出パターン発見法を開発し(DBS研, ACM ICUIMC'09)、情報処理学会データベースシステム研究会学生奨励賞を受賞した(大谷, 2008.06.20)。

不完全データに対する柔軟な半構造マイニング手法として、距離できまる近接性のみを考慮した近接パターンマイニングの枠組みを研究した。具体的には、サイズ限定木構造パターンのクラスに対するパターン列挙問題を考察し、与えられた木構造データにおいて、このクラスのパターンを定数遅延時間列挙する超高速発見手法の開発に成功した(DEIM'12, COMP研'12)。この成果は、2012年の組合せと計算会議(COCOON2012)で発表予定であり、また信学会 DEIM2012 若手奨励賞を受賞した(和佐他, 2011.3)。

さらに、実世界の大规模位置データに対する半構造マイニングの拡張として、従来の離散的な系列マイニングを連続的な実数座標をもつ軌跡データに拡張し、GPSデータに対する閉パターンマイニングの研究を行った。

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

深さ優先探索に基づくあいまい性をゆるした頻出パターンマイニング手法を開発し、第12回環太平洋データマイニング会議で発表した(PAKDD'08)。木構造マイニングへの確率スキーマの導入として、根付き木構造上の

階層ベイズモデルを提案し、効率よいEMアルゴリズムを開発し、実証実験として、インフルエンザウイルスの巡回パターンの解析に適用した(GIW'09, SIGBIO'10)。本研究は、情報処理学会平成22年度山下記念研究賞(柳橋, 2010.05.22)を受賞した。

系列マイニングへの確率スキーマの導入に関しては、確率的刈込み接尾辞木を用いた系列予測手法を開発しスパムフィルタリングやウェブマイニングの実証実験を行った(DS'09, Trans. JSAI'10)。さらに、半構造データ系列の索引として、可変長符号化文字列に対する全文テキスト索引の研究を行い、符号化接尾辞木の線形時間構築アルゴリズムを与えた(CPM'11)。さらに、近接パターンマイニングアルゴリズムの統計学習への適用、および、根付き木構造上の確率モデルの学習による、半構造マイニングデータの自動生成への研究検討を行った。

(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)。

これまでの最右拡張法に基づく高速木構造マイニングおよび、極大パターンマイニング手法 LCM の理論の全体像について、第12回環太平洋データマイニング国際会議の基調講演で紹介した(PAKDD'08)。

さらに、これらの技術を一般化し、系列、木、グラフに対する頻出閉集合パターン発見問題を、アクセス可能という性質をみたす集合族に対する閉集合パターン発見問題として定式化し、多項式遅延多項式領域アルゴリズムの一般的構成法を与えた。この理論に基づいて、未解決のいくつかの半構造パターンクラスに対する統一的な低メモリで高速なアルゴリズムを開発した(SDM'09, ISCSIS'10)。

(4) 大规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(喜田・伊藤・湊・有村)。

ネットワーク上の多様な半構造データからの半自動知識関係を実現するための検索と抽出技術を研究開発した。とくに、膨大なウェブ上のテキストデータに対する低メモリかつ高速な処理のために、データを圧縮したまま高速に処理する圧縮パターン照合技術に焦点を当てた研究を行った。

具体的には、圧縮パターン照合に適した可変長情報源固定長符号(VF符号)の研究開発を行い、良く知られた半構造データ索引である接尾辞木に基づくST-VF符号化方式を提案し、この分野の主要会議であるIEEEデータ圧縮国際会議で発表した(IEEE DCC'09)。さらに、種々のVF符号に対して、メモリ効率解析や(IEEE DCC'10)、圧縮パターン照合の性能解析(IEEE DCC'11)、機械学習のアイデ

ィアを用いた高圧縮率をもつ解析木の訓練手法 (SPIRE' 10, JIP' 12) を研究開発した。

また、木構造データの最小汎化技術を用いた知識連携技術を研究し、ウェブから対話的情報抽出手法を開発し (iDB フォーラム' 08)、情報処理学会データベースシステム研究会学生奨励賞を受賞した (筒井, 2008.06.20)

さらに、近年のネットワーク不正侵入検出などの大規模ストリーム系列検索においては、複雑なパターンに対して、ソフトウェアとハードウェアを結合した高速な検索技術の実装法が重要である。そのために、膨大なウェブ上の半構造系列データに対して高機能かつ高速な情報抽出を実現するために、多次元データストリームのための高速半構造パターン照合技術を開発した (ALSIP' 08)。

さらに、複合イベントストリーム処理 (CEP, ESP) に用いられる正規表現の部分クラスに対して、ビット演算と算術演算を直接利用するビット並列計算技術に基づく高速かつ低メモリのパターン照合手法を開発した (SPIRE' 10)。

さらに、これを木構造パターン (IWOCA' 10, JDA' 12) に拡張した。他方で、カウンタ付き NFA を用いた窓付き正規表現照合に拡張し (DEIM' 12)、信学会 DEIM2012 若手奨励賞 (細川他, 2011.3) を受賞した。

さらに、これらの技術の再構成可能ハードウェア (FPGA) 上の並列ハードウェア実装を行い、拡張文字列パターンのクラスに対して世界最高速のパターン照合速度を達成した (IEEE FPT' 10, SASIMI' 10, APSIPA' 11)。

(5) ZDD に基づく高速な大規模知識索引技術の研究開発 (湊・喜田・有村)。

分担者の湊が開発した大規模論理最適化技術であるゼロサプレス二部決定グラフ (ZDD) に基づく、大規模知識索引技術の研究開発を行った。

2008 年には、分担者の宇野が開発し、世界最速の閉集合パターン発見技術として知られている LCM アルゴリズムを、湊が開発した ZDD に基づく知識索引技術と結合し、与えられたトランザクションデータベースから、出現するすべての頻出集合パターンすべてを含む圧縮知識索引を高速に直接構築する手法 LCMoverZDD を開発し、大規模実データ上で評価実験を行った (PAKDD' 08)。

この知識索引技術に基づき、データベース系列から、与えられた生起パターン表現をもつ頻出アイテム集合の発見手法 (SDM' 10) や、包含パターン発見アルゴリズム (IEEE CIT' 08) の開発、その性能評価 (LLLL' 09) を行った。

項目 (2) との連携として、ZDD を用いたベイジアンネットワークモデルの超高速確率計算法を研究開発した (SIG-FPAI72, DMSS' 09)。

さらに、BDD の出力依存計算量に関する長年の未解決問題を否定的に解決した (IPL, 2012)。ZDD を系列データに拡張した系列 BDD (Loekit, Bailey) の計算論的特性の解析を行い、種々の高速演算を開発した。

さらに、ZDD に基づく高速な大規模知識索引技術の成果について招待講演を行った (IWMSP' 08)。さらに、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) 実装と評価実験 (全員)。

開発した知識発見技術について、研究開発と並行して、実装を行い、最適化を行った。さらに、知識基盤形成支援システムのプロトタイプ構築と評価実験を行い、これらに基づいて、要素技術のさらなる最適化を行った。なお、これらの実装と計算機実験のために、研究費からサーバと開発用 PC を購入した。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線) 末尾の 1 (0) は査読の有 (無) を示す。

[雑誌論文] (計 22 件)

1. Y. Kaneta, S. Yoshizawa, S. Minato, H. Arimura, and Y. Miyayaga: "A Dynamically Reconfigurable FPGA-based Pattern Matching Hardware for Subclasses of Regular Expressions," *IEICE Trans. Inf. and Syst.*, Vol.D (2012), 1
2. S. Yoshida, T. Uemura, T. Kida, T. Asai, and Seishi Okamoto: "Improving Parse Trees for Efficient Variable-to-Fixed Length Codes," *Journal of Information Processing*, Vol. 20, No. 1, 238-249 (2012), 1
3. R. Yoshinaka, J. Kawahara, S. Denzumi, H. Arimura, and S. Minato, "Counterexamples to the long-standing conjecture on the complexity of BDD binary operations," *Information Processing Letters*, Vol.112, Issue 16, 636-640 (2012), 1
4. Y. Kaneta, H. Arimura, R. Raman: "Faster Bit-Parallel Algorithms for Unordered Pseudo-tree Matching and Tree Homeomorphism," *Journal of Discrete Algorithms* (2011), 1
5. Y. Okamoto, T. Uno: "A polynomial-time-delay and polynomial-space algorithm for enumeration problems in multi-criteria optimization," *European Journal of Operational Research* 210 (1), 48-56 (2011), 1
6. K. Hirata, Y. Yamamoto, T. Kuboyama: "Improved MAX SNP-Hard Results for Finding an Edit Distance between Unordered

- Trees," Proc. CPM 2011, LNCS, Springer, Vol.6661, 402-415, 2011, 1
7. T. Uemura and H. Arimura: "Sparse and Truncated Suffix Trees on Variable-Length Codes," Proc. CPM2011, LNCS, Springer, Vol.6661, Springer, 246-260, 2011, 1
  8. T. Katoh, H. Arimura, K. Hirata: "An Efficient Depth-first Search Algorithm for Extracting Frequent Diamond Episodes from Event Sequences" IPSJ Online Transactions, TOD, Vol.3. 1-12 (2010), 1
  9. T. Uemura, D. Ikeda, T. Kida, H. Arimura: "Unsupervised Spam Detection by Document Probability Estimation with Maximal Overlap Method" Transactions of the Japanese Society for Artificial Intelligence Vol.26. 297-306 (2010), 1
  10. 喜田拓也: "分節木と共用文字列で表現される符号上での効率良い圧縮照合アルゴリズム" 電子情報通信学会和文論文誌 Vol.J93-D. 733-741 (2010), 1
  11. Y. Kaneta, H. Arimura: "Faster Bit-Parallel Algorithms for Unordered Pseudo-tree Matching and Tree Homeomorphism" Proc. IWOC2010, LNCS, Springer, Vol.6460. 68-81 (2010), 1
  12. Y. Kaneta, S. Minato, H. Arimura: "Fast Bit-Parallel Matching for Network and Regular Expressions" Proc. SPIRE'10, LNCS, Vol.6393. 372-384 (2010), 1
  13. T. Katoh, H. Arimura, K. Hirata: "An Efficient Depth-First Search Algorithm for Extracting Frequent Diamond Episodes from Event Sequences" IPSJ Trans. on Databases Vol.2, No.4. 1-12 (2009), 1
  14. 喜田拓也: "STVF 符号:頻度刈り込み接尾辞木を用いた効率良いVF符号化" 日本データベース学会論文誌 DBSJ Journal Vol.8. 125-130 (2009), 1
  15. T. Katoh, H. Arimura, K. Hirata: "A Polynomial-delay Polynomial-space Algorithm for Extracting Frequent Diamond Episodes from Event Sequences" Proc. PAKDD2009, LNAI, Springer, Vol.5476. 172-183 (2009), 1
  16. T. Katoh, H. Arimura, K. Hirata: "Mining Frequent Bipartite Episode from Event Sequences" Discovery Science 2009, LNCS, Springer, Vol.5808. 136-151 (2009), 1
  17. 上村, 喜田, 有村: "ウェブ閲覧における効率的なキーワード抽出とその利用" 情報処理学会撤誌 : データベース(TOD) 38. 49-60 (2008), 1
  18. 上村, 喜田, 有村: "プロパティ接尾辞木のオフライン線形時間構築アルゴリズム" 電子情報通信学会論文誌 J91-D. 595-607 (2008), 1
  19. T. Uno and H. Arimura: "Ambiguous Frequent Itemset Mining and Polynomial Delay Enumeration" LNCS, Springer, Vol.5012. 357-368 (2008), 1
  20. S. Minato, T. Uno, and H. Arimura: "LCM over ZBDDs : Fast Generation of Very Large-Scale Frequent Itemsets Using a Compact Graph-Based Representation" LNCS, Springer, Vol.5012. 234-246 (2008), 1
  21. T. Asai, S. Okamoto, and H. Arimura: "An Adaptive Algorithm for Splitting Large Sets of Strings and Its Application to Efficient External Sorting" LNCS, Springer, Vol.5433. 13-25 (2008), 1
  22. T. Kida, T. Saito, and H. Arimura: "Flexible Framework for Time-Series Pattern Matching over Multi-Dimension Data Stream" LNCS, Springer, vol.5433, 1-12 (2008), 1
- [学会発表] (計 25 件)
1. Kunihiko Wasa, Takeaki Uno, and Hiroki Arimura: "Constant Time Enumeration of Bounded-Size Subtrees in Trees and Its Application," Proc. COCOON'12, LNCS, Springer, 2012. (to appear)
  2. 細川 威樹, 金田 悠作, 有村 博紀: "複合イベント系列処理のための効率良い窓付きパターン照合アルゴリズム, DEIM2012, B8-1, 2012.
  3. 和佐 州洋, 金田 悠作, 宇野 毅明, 有村 博紀: "半構造データに対する効率良い近接パターン照合アルゴリズム," DEIM2012, 2012.
  4. S. Denzumi, Ryo Yoshinaka, H. Arimura and Shin-Ichi Minato: "Notes on Sequence Binary Decision Diagrams: Relationship to Acyclic Automata and Complexities of Binary Set Operations," Proc. PSC'11, Czech Tech. Univ., Prague, 2011.
  5. S. Denzumi, H. Arimura, S. Minato: "Implementation of Sequence BDDs in Erlang," Proc ACM ERLANG, 90-91, ACM, Sep. 2011.
  6. Y. Kaneta, S. Yoshizawa, S. Minato, and H. Arimura: "High-speed String and Regular Expression Matching on FPGA," Proc APSIPA ASC 2011, Oct. 2011.
  7. H. Arimura: "Efficient Algorithms for Discovering Frequent and Maximal Substructures from Large Semistructured Data," Proc. ISICIS'10, LNEE, Vol.62, Springer, 353-358, (2010). London, UK.
  8. S. Yoshida, T. Kida: "An Efficient Algorithm for Almost Instantaneous VF Code Using Multiplexed Parse Tree" Proc. DCC'10, IEEE, 219-228 (2010), 1
  9. Y. Kaneta, S. Minato, H. Arimura, 他 2 名:

- "Dynamic Reconfigurable Bit-Parallel Architecture for Large-Scale Regular Expression Matching" Proc. IEEE FPT'10, IEEE, 21-28 (2010), 1
10. S. Minato, T. Uno: "Frequentness- Transition Queries for Distinctive Pattern Mining from Time-Segmented Databases" Proc. SDM'10, SIAM, 339-349 (2010), 1
  11. S. Yoshida, T. Kida: "On Performance of Compressed Pattern Matching on VF Codes" Proc. DCC'11, IEEE, 486 (2011), 1
  12. T. Katoh, K. Hirata, H. Arimura, 他 2 名: "Aligned Bipartite Episodes Between the Genera of Bacteria" Proc. 2010 CME'10, IEEE/ICME, 193-197 (2010), 1
  13. Y. Kaneta, S. Minato, H. Arimura: "Efficient Multiple Regular Expression Matching on FPGAs based on Extended SHIFT-AND Method," Proc. SASIMI'10, (2010). Taipei.
  14. H. Arimura, T. Uno: "Polynomial-Delay and Polynomial-Space Algorithms for Mining Closed sequences, Graphs, and Pictures in Accessible Set Systems" Proc. SDM'09, SIAM, 1087-1098, 2009, 1
  15. S. Minato: "A Fast Algorithm for Cofactor Implication Checking and Its Application for Knowledge Discovery," Proc. IEEE CIT 2008, 2008/07/8, Sydney, Australia.
  16. Haruya Iwasaki, Elsa Loekito, S. Minato, James Bailey: "Comparison of ZDD-Vectors and WZDDs for Frequent Pattern Mining," Proc. LLLL'09, (2009), Kyoto.
  17. T. Katoh, H. Arimura, Koichi Hirata: "Mining Frequent k-Partite Episodes from Event Sequences," Proc. LLLL'09, (2009), Kyoto.
  18. D. Tokoro, Kiyoharu Hama guchi, Toshinobu Kashiwabara, S. Minato: "Monolithic and Partial Compilation Methods for Probabilistic Inference of Bayesian Networks using ZBDDs," DMSS'09. (2009. 京大会館(京都府)).
  19. Fumiaki Yanagihashi, Kimihito Ito, H. Arimura: "Inference of Geographic Transmission Probability of In uenza Viruses from a Large Phylogenetic Tree," Proc. GIW '09 (2009), Pacifico Yokohama.
  20. 柳橋史成, 伊藤公人, 有村博紀: "木の最適ラベリング問題とその進化系統樹への応用," 第 16 回 SIG BIO, (2009). 中央大学後楽園キャンパス.
  21. H. Arimura: "Efficient Algorithms for Mining Frequent and Closed Patterns from Semi-structured Data (invited talk)," PAKDD'08, Lecture Notes in Computer Science (LNCS), Springer, Vol.5012. 2-13 (2008), 0
  22. S. Minato: "Efficient Method of Large-Scale

Data Analysis Using BDD-based Data Compression (Invited Talk)," In Proc. IWMMSP'08, 2008, 0.

23. S. Minato: "ZBDD-Based Probabilistic Computation and Application for System Analysis" In Proc. of Workshop of SoC Design Methodologies, 2008, National Tsing Hua University, China.
24. 筒井淳平, 有村博紀: "XQUBE : 具体例と演示からの XQuery 問合せ構築のための視覚言語," iDB フォーラム 2008, (2008). 福島, 飯坂ホテル聚楽.
25. 大谷英行, 喜田拓也, 宇野毅明, 有村博紀: "極小出現区間を用いた頻出エピソード発見アルゴリズムの高速化," 情報処理学会 jDB フォーラム DBS 研 (2008). 小樽市民会館.

[図書] (計 1 件)

Thomas Zeugmann, 湊真一, 大久保好章: "英語で学ぶ計算理論," コロナ社. 222, 2008.

[その他]

ホームページ等

<http://www-ikn.ist.hokudai.ac.jp/~arim/>

## 6. 研究組織

### (1) 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  
研究者番号 : 20222763

### (2) 研究分担者

喜田 拓也 (KIDA TAKUYA)

北海道大学・大学院情報科学研究科・准教授  
研究者番号 : 70343316

湊 真一 (MINATO SHIN-ICHI)

北海道大学・大学院情報科学研究科・准教授  
研究者番号 : 10374612

伊藤 公人 (ITO KIMIHITO)

北海道大学・人獣共通感染症リサーチセンター・准教授

研究者番号 : 60396314

宇野 毅明 (UNO TAKEAKI)

国立情報学研究所・情報学プリンシプル研究系・准教授

研究者番号 : 00302977

平田 耕一 (HIRATA KOUICHI)

九州工業大学・情報工学研究院・准教授

研究者番号 : 20274558