

VII. 計算情報学研究部門

VII-1. データ基盤分野

1. メンバ

教授 北川博之

准教授 天笠俊之

講師 川島英之

2. 概要

e サイエンスにおいて、大規模データの管理や活用は極めて重要な課題となっている。計算情報学研究部門データ基盤分野は、データ工学関連分野の研究開発を担当している。具体的には、異種データベースや多様な情報源を統合的に扱うための情報統合基盤技術、データ中に埋もれた知識や規則を発見するためのデータマイニング・知識発見技術、インターネット環境において様々なデータを統合的に扱うための XML 関連技術の研究を継続して行った。また、センター内の地球環境研究部門や素粒子物理研究部門や産業技術総合研究所と連携して、計算科学の各分野における応用的な研究を推進した。

3. 研究成果

2011 年度の主要な研究成果について説明する。

【1】情報統合基盤技術

(主な研究費: 科研費基盤研究 (A), 科研費特定領域研究, 科研費若手研究 (B))
分散した異種情報源を統合的に扱うための基盤技術・システム・応用の研究開発を行った。特に、従来型のデータベースや Web 等のみではなく、センサ、位置情報源等の連続的に情報を提供するストリーム情報源をも対象とすることができるストリーム処理基盤システムにおけるデータ処理技術を中心に研究開発を行った。

(1) 高信頼化ストリーム処理 (北川, 川島) [P22]

センサデータなどのストリームデータに対する問合せ要求が増大し、それらを実現するストリーム処理システムが研究開発されている。地理的に離れた情報源の統合や負荷分散を実現するために、ストリーム処理システムを分散配置した上で協調動作させる分散ストリーム処理環境が注目されている。分散環境では、中継ノードが停止することでシステム全体が停止してしまうという問題がある。この問題に対して、我々は分散環境において高信頼化を実現する Semi-Active Standby 方式を提案してきた。これは、各分散ノード間の通信にバッチ処理を導入することで、既存手法である Active Standby 方式, Upstream Backup 方式を一般化し、リカバリ時間とバンド幅オーバーヘッドを適用環境に合

わせて調節することを可能とする．本研究では Semi-Active Standby 方式における ack パラメタ調整を自動化する方式について検討を行った．そして様々な設定において評価実験を行い，本方式の動作特性を検証した．

(2) ストリームデータの永続化手法 (川島, 北川) [C11]

本研究では各種デバイスから自律的，かつ継続的に発信されるストリームデータを永続化する手法を検討した．永続化とは，停止故障が生じてもデータを失わないデバイスである永続的デバイスへデータを書き込むことによって，ストリームデータを永続的に保存する技術である．高頻度で配信されるストリームデータを永続化するには，ディスクへの書き込み処理がボトルネックになる．この問題を解決するための永続化手法を検討した．第一の提案は，問合せ処理中に出現する複数の選択演算子を併合することである．これにより永続化の対象となるデータ量を削減することが可能になる．第二の提案は，問合せ処理木の途中で，ストリームデータを中間領域と呼ばれるファイルへ一時的にシーケンシャルに書き込む事である．処理木とは，問合せの代数的表現であり，ストリーム処理システム内部でデータ処理を記述するために使用される構造である．この提案によりランダムアクセスを無くし，書き込み性能を向上させる．ただし中間領域のデータはユーザからアクセス不能である為，中間領域のデータをユーザからアクセス可能にするため，それらを遅延して読み出し，データ領域へ書き込むことが必要になる．第三の提案は，中間領域からの読み出しが書き込みに干渉しないように，データ領域への書き込み処理を制御することである．提案手法によりストリームデータの永続化が効率化されることを実験的に示した．

(3) メディアストリームとタプルストリームの統合的処理 (川島, 北川)

[P20]

これまでストリームデータを処理するための枠組みとして，ストリーム処理システム(Stream Processing System) という枠組みが研究，開発されてきた．しかし，既存のストリーム処理システムはタプルストリームの処理を前提としているために，画像，動画のようなメディアストリームの処理を行うことが難しい．そこで本研究ではメディアストリームを管理するメディアサーバおよびメディアデータ要求インタフェースを開発した．なお，メディアストリームの情報源としては Kinect を対象とした．本研究の貢献を下記に示す．

✓ メディアサーバ

メディアサーバはメディアストリームを受信すると，それを保持する．メモリが不足した場合にはデータを一部の二次記憶へ移動する．メディアサーバはデータ配信要求を受け取ると，該当するメディアストリームを配信する．メディアサーバは複数のクライアントからの同時接続を許可する．

✓ メディア要求インタフェース

メディア要求インタフェースは、あるタプルストリーム処理系からのタプルを受信した時に、メディアストリームを取得するインタフェースを提供する。ある情報源からのタプルが到着した時に処理木を動作させる仕組みについては、すでに StreamSpinner において MASTER 節が提案されている。しかし、あるタプルストリーム処理系からのタプルを受信したときに、別のストリーム処理系からのデータを受信するスキームは我々が知る限り存在しない。本研究の提案と MASTER 節の違いは処理レベルの高さにある。MASTER 節は 1 つの処理木におけるイベント駆動スキームであるが、本研究は異なるストリーム処理系におけるイベント駆動スキームである点が異なる。

✓ Kinect データ取得システム

本研究で扱うメディアデータ情報源として扱う Kinect は様々な種類のデータを取得可能である。Kinect をサーバ化した研究には KineX[9] が存在する。しかし KineX は Windows でしか動作しない。そこで本研究では Linux 環境で動作する Kinect データサーバである Konnect を実装した。

【2】データマイニング・知識発見技術

(主な研究費：科研費基盤研究 (A) , 科研費若手研究 (B))

多様な情報源に対する外れ値検出, ソーシャルネットワーク分析, 種々のデータマイニング・知識発見技術の研究開発を進めた。以下では、主なものについて述べる。

(1) マイクロブログの分析 (天笠, 北川) [J5, C2, P11]

Twitter におけるメッセージはツイートと呼ばれ、ツイートを投稿することをポストという。Twitter ユーザはツイートに外部の Web ページの URL を含め、他のユーザと共有することが多い。これは、自らが投稿したブログなどの外部ページの URL をツイートに含めることで他のユーザに知らせたり、閲覧して面白い、参考になると感じた Web ページを他のユーザと共有したりすることによる。予備調査によると、ポストされたツイートのうち約 14 % は URL を含んでいることが明らかになった。共有された Web ページは、生成されて間もないページであることが多いと報告されている。本研究では、生成されて間もない Web ページを鮮度の高いページと呼ぶ。したがって、Twitter からは鮮度の高い Web ページを得ることができると考えられる。しかし、Twitter 上で共有される Web ページは膨大であり、そのトピックも多岐にわたっているため、その中から自らが得たいものを得るのは困難である。そこで本研究では、膨大な Web ページの中から重要なもののみをユーザの興味に合わせて推薦する手法を提案した。詳細は下記に述べる。

まず、ある Web ページがどのような専門性を持つユーザに共有されたかによって、その Web ページの重要度を推定した。ユーザの専門性とは、ユーザが主にどのようなトピックに関しての情報を発信するかという属性である。例えば、スポーツに関する情報をよく発信する多くのユーザがある Web ページを共有

していれば、その Web ページは重要であり、かつスポーツに関するものであると考えられる。次に、ユーザの興味を推定し、その興味に見合った Web ページを推薦した。ユーザの興味は、そのユーザがフォローするユーザ群の専門性によって推定した。例えば、スポーツに関して専門性の高いユーザを多くフォローするユーザは、スポーツに興味を持っていると考えられる。実験により提案手法の有用性を確認した。

(2) 再生リストを利用した Web 動画クラスタリング (北川) [C10, P7]

近年複数の動画共有サイトで普及している機能として、「再生リスト」がある。再生リストはユーザが任意に作ることの出来る動画のリストで、ユーザは再生リストに登録されている動画に素早くアクセスする事ができる。一人のユーザが複数の再生リストを作成可能なので、音楽動画を集めた再生リストやスポーツに関する動画を集めた再生リストなど、再生リストを利用して頻繁にアクセスする動画をカテゴリ化することが出来る。その場合、同一再生リストに含まれる動画間にはなんらかの類似性があると考えられる。我々はこの「再生リスト」に着目し、再生リスト情報に基づいて動画間の類似度計算を行う。そして画像や音声、テキスト情報に依存しない高精度な動画クラスタリング手法を提案する。本研究では最初に、予備実験によって同一再生リストに含まれる動画間に類似性があるかどうかを確認した。次に再生リスト情報に基づく動画クラスタリング手法を提案した。実験によって提案手法の効果を定量的に示した。

(3) 距離索引を用いた逆最遠傍問合せに関する効率的な検索手法 (北川) [c9, P3]

近年、利用者の多様な情報要求に対応するため、多次元データに対するさまざまな問合せ処理が提案されている。k 近傍問合せ (k-Nearest Neighbor Query)、範囲問合せ (Range Query)、逆最近傍問合せ (Reverse Nearest Neighbor Query) などはその例である。しかし、問合せ点から遠傍に存在するデータの検索、特に、逆最遠傍 (Reverse Furthest Neighbors; RFN) の問合せ処理あまり注目されていなかった。

データセット O 及びクエリ q を与えた際に、逆最遠傍検索は、 O の中から q を最遠傍とする全てのオブジェクトを求める問題である。本研究では、逆最遠傍 (Reverse Furthest Neighbors, RFN) 探索問題に対して、 R 木構造を基に特化した空間索引と、これを用いた検索アルゴリズムを提案した。先行研究では、従来手法の高い計算量を大幅に削減でき、任意位置のクエリ q に適用可能な検索手法を提案したが、効率的に十分ではなく、多次元データにも適応しにくいという問題があった。本研究では、ピボットセットを選出して、同一ピボットを最遠傍とするデータをまとめて R 木に格納し、最遠傍距離を埋め込めるように特化した空間索引、およびこの空間索引を用いた RFN 検索アルゴリズムを提案した。さらに、多次元への拡張を検討し、合成データと実データを用いて提案手法の効率とスケーラビリティを検証した。

(4) GPU によるデータマイニング処理の高性能化 (天笠, 北川) [C1, C3, P2]

GPU は数百以上の SIMD (Single Instruction, Multiple Data) コアを持つため, CPU と比較して, 極めて並列度の高いデータ処理を行なうことができる. この SIMD コアは, 一つの命令を複数のデータに対して, 並列に実行していく処理に適したアーキテクチャとなっている. そのため, GPU はその性質上, 複雑な条件分岐を含む処理には向いていない. そこで, GPGPU を用いる際には, GPU に適したアルゴリズムの開発が重要な課題の一つとなる. 本研究では, 不確実データベースからの頻出アイテム集合マイニングの高速化のために, GPGPU を利用した手法を提案した. 提案手法は, Sun らの提案した pApriori アルゴリズムを元に, GPU 上での並列実行により処理の高速化を図った. さらに, pApriori アルゴリズムとの比較実験により, 提案手法の性能を評価した.

【3】XML・Web プログラミング

(主な研究費: 科研費基盤研究 (A), 科研費若手研究 (B))

Web 上の標準データフォーマットとして利用されている XML (Extensible Markup Language), メタデータ記述の枠組みである RDF (Resource Description Framework) に関する研究を行うとともに, オープンな環境でのデータ利用に書かせないデータプライバシーの研究を行った.

(1) XML データに対するファセット検索 (天笠, 北川) [C8, P4]

グラフは複雑なデータを表現可能なデータ構造である. 一方で, ファセット探索は探索的な検索を行うための有用な手法として広く利用されている. グラフデータに対してファセット探索を行う際には, 検索対象 (ノードや部分グラフ) とプロパティ (特徴を表す属性) が予め定義されている必要がある. しかしながら, グラフデータ中の検索対象や検索対象のプロパティを予め指定することは容易ではない. これに対して本研究では, グラフデータに対するファセット探索を行うための検索対象の抽出方法について議論する. 本研究ではグラフ中の頻出パターンに着目することで検索対象の抽出を行う枠組みを提案する. また, 本研究では抽出した検索対象やファセットを効率的に格納する方法についても議論した.

(2) 省電力を考慮した XML ストリーム処理 (天笠, 北川) [P10]

XML は多様なデータを柔軟に表現できるため, 多くの分野でデータの記述, 交換のために利用されている. Web サービスやセンサネットワークにおいては, データの送受信のために XML フォーマットを利用することが多い. このとき, XML データは, 情報源から継続して送信される XML ストリームとみなすことができ, XML ストリームに対する問合せやフィルタリング処理に関する研究が数多く行なわれてきた. 一方, 地球温暖化や震災による電力不足をきっかけとして, 情報システムにおいて電力消費の問題は特に注目されている. 前述の XML ストリ

ーム処理系は、常時オンで処理を続けるため、電力削減の効果が期待できる。

このため、本研究では XML ストリーム処理における省電力を考慮した問合せ処理方式を検討した。本研究の着眼点は、1) XML ストリームにおいて、問合せ結果として必要な部分は多くの場合ごく一部であること、2) XML ストリーム処理では、文字列の字句解析、構文解析に多くの計算量が割かれていることである。このため、実際の問合せ結果として必要でない部分については、構文解析等の処理を省くことで、省電力化を達成した。具体的には、XML のパーサと問合せ処理部を拡張し、相互に連携をする機構を開発することによって、問合せ結果になりえない XML データについては、構文解析の処理を省くことを可能にした。この成果が評価され、DEIM フォーラム 2012 において、学生奨励賞を受賞した。

(3) Lined Data のスキーマ情報推定 (天笠, 北川) [P6]

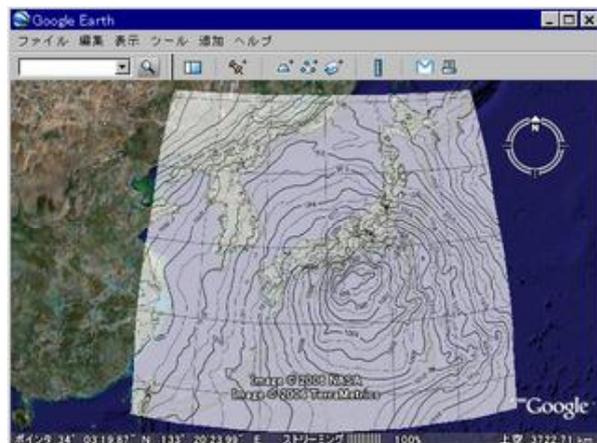
Web 上でのデータの相互運用性を改善するための枠組みとして、Linked Data が注目されている。Linked Data は、RDF (Resource Description Framework) を用いながら、実際に解決可能な URL を用いることで、データ同士を相互にリンクする試みであり、統計データや観測データなどが実際に Linked Data を利用して公開されつつある。ところが、提供されるデータのほとんどに、スキーマ情報が付随していないという問題がある。スキーマ情報は、推論や検索処理の際に有用であり、スキーマ情報を導出することは極めて有用である。そこで、本研究では、スキーマのない Linked Data に対してスキーマ情報を推測する手法を提案した。これは、共通するリソースが似た URL のパターンを共有する点に着目した手法である。実際に、クラス間の継承関係 (`rdfs:subClassOf`)、プロパティの定義域 (`rdfs:domain`) と値域 (`rdfs:range`)、同値関係 (`owl:sameAs`) が一定の精度で推定可能であることを示した。

【4】科学分野におけるデータベース応用

(主な研究費：科研費若手研究 (B))

(1) GPV/JMA アーカイブ (天笠, 北川)

地球環境研究部門と共同で、気象庁気象予報データベース「GPV/JMA アーカイブ」(<http://gpvjma.ccs.hpcc.jp>) の開発、および管理、運用を行っている。GPV/JMA アーカイブは、気象庁が公開している気象予報グリッドデータ (GPV データ) を蓄積するとともに、外部登録ユーザへのデータを提供することを目的としている。GPV/JMA アーカイブで提供している



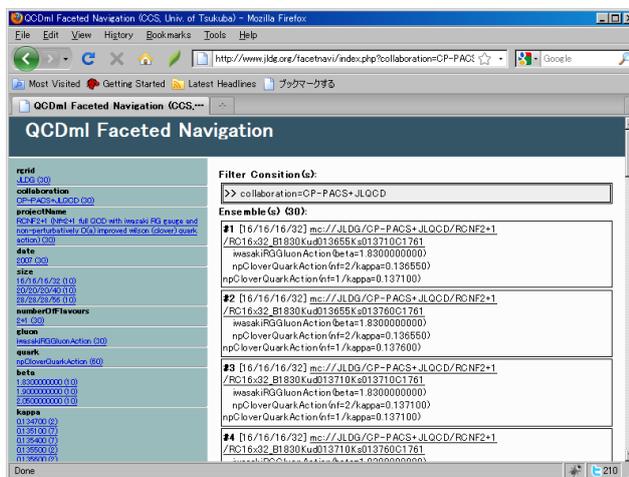
データは、全球モデル、メソスケールモデル、リージョナルスケールモデル、週間アンサンブル、月間アンサンブル、季節アンサンブルの 6 種類である。さらに、これらのグリッドデータに加えて、数値データを元に作図した天気図を公開するとともに、天気図の閲覧性の向上するため、GoogleEarth 上に天気図をマップするための KML ファイルの公開サービスも行っている。

(2) 大規模衛星センサデータからのイベント検出 (川島, 北川) [P9]

GEO Grid がアーカイブしている地球観測データの中には TIR(Thermal Infrared Radiometer)と呼ばれる熱に関するデータがある。熱放射特性を利用して鉱物資源の判別や大気、地表面、海面の状態を観測することを主な目的としている。本研究では、製鉄所や火災を起きている場所など、周辺の領域より著しく温度が高い領域の検出を目的として、TIR データに基づくホットスポットの検出手法を提案した。提案手法は閾値に基づく手法と統計値に基づく手法の 2 つがある。提案手法により、製鉄所、森林火災を検出できた。

(3) 格子 QCD アンサンブル XML のファセット検索 (天笠, 北川)

ILDG (International Lattice Data Grid) では、格子 QCD 配意データのメタデータとして XML が用いられている。世界中の地域グリッドで公開されている配意データを検索するため、利用性の高いインタフェースの開発が望まれていた。このため我々は、QCDml のためのファセット検索インタフェースを設計、実装を行った。ファセット検索とは、検索対象オブジェクトの集合を効率よく探索するための手法である。オブジェクトは、あらかじめファセットと呼ばれるいくつかの独立したカテゴリ毎に分類されている。各カテゴリ(ファセット)において、オブジェクトは着目する属性の値毎にグルーピングされており、その値がリスト表示されている。利用者はファセットに含まれる具体的な値を選択することで、オブジェクトの絞り込みを行い、探索を行う。XML データに対してファセット検索を適用するため、XML は半構構性を考慮したファセットの抽出および QCDml におけるファセットの検討を行い、実際にシステムを構築した。



4. 研究業績

(1) 学術雑誌論文

J1. 三好健文, 寺田裕太, 川島英之, 吉永努, “ストリーム処理エンジン向け動

的再構成可能プロセッサアーキテクチャの設計”, 情報処理学会論文誌：データベース, Vol. 4, No. 2 (TOD50), pp. 35 - 51, 2011.

- J2. Cui Zhu, Hiroyuki Kitagawa, Spiros Papadimitriou, and Christos Faloutsos, "Outlier Detection by Example", *Journal of Intelligent Information Systems*, Vol. 36, No. 2, pp. 217-247, 2011.
- J3. 三好健文, 寺田裕太, 川島英之, 吉永努, “ウィンドウ結合演算子の FPGA による実現”, 電子情報通信学会論文誌通信ソサイエティスマートな社会を支えるインターネットアーキテクチャ特集号, J94-B(10), 1313-1322, 2011-10-01
- J4. 猿渡俊介, 川島英之, 高木潤一郎, 倉田成人, 森川博之, “センサデータベースマネージャにおける問合せ処理とデータ圧縮の同時最適化”, 情報処理学会論文誌「新たな展開を迎える ITS、モバイル通信とユビキタスコンピューティング」特集号, 53(1), 320-335, 2012-01-15
- J5. 山口祐人, 天笠俊之, 高橋翼, 北川博之, “情報伝搬を考慮したグラフ分析による Twitter ユーザランキング手法”, 情報処理学会論文誌データベース, Vol. 4, No. 2 (TOD50), pp. 142-157, 2011.

(2) 国際会議論文

- C1. Eli Koffi Kouassi, Toshiyuki Amagasa, and Hiroyuki Kitagawa, "Efficient Probabilistic Latent Semantic Indexing using Graphics Processing Unit", *Proc. International Conference on Computational Science (ICCS 2011)*, Singapore, pp. 382-391, June 1-3, 2011.
- C2. Yuto Yamaguchi, Toshiyuki Amagasa, and Hiroyuki Kitagawa, "Tag-based User Topic Discovery using Twitter Lists" *Proc. International Conference on Advances in Social Network Analysis and Mining (ASONAM 2011)*, Kaohsiung City, Taiwan, pp.13-20, July 25-27, 2011.
- C3. Yusuke Kozawa, Toshiyuki Amagasa, and Hiroyuki Kitagawa, "Fast Frequent Itemset Mining from Uncertain Database Using GPGPU", *Proc. Fifth International VLDB Workshop on Management of Uncertain Data (MUD 2011)*, Seattle, Washington, pp. 17-24, August 29, 2011.
- C4. Kazutaka Furuse, Hiroaki Ohmura, Hanxiong Chen, and Hiroyuki Kitagawa, "An Extended Method for Finding Related Web Pages with Focused Crawling Techniques", *Proc. 15th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2011)*, Kaiserslautern, Germany, pp. 21-30, September 12-14, 2011.
- C5. Masafumi Oyamada, Hideyuki Kawashima, and Hiroyuki Kitagawa, "Efficient Invocation of Transaction Sequences Triggered by Data Streams", *The 2nd International Workshop on Streaming Media Delivery and Management Systems (SMDMS 2011)*, Proceedings of

- 3PGCIC, Barcelona, Spain, October 26-28, 2011.
- C6. Tsubasa Takahashi, Hideyuki Kawashima, and Hiroyuki Kitagawa, "A Video Manager for Relational Stream Processing Systems", The 2nd International Workshop on Streaming Media Delivery and Management Systems (SMDMS 2011), Proceedings of 3PGCIC, Barcelona, Spain, October 26-28, 2011.
- C7. Yasin Oge, Takefumi Miyoshi, Hideyuki Kawashima and Tsutomu Yoshinaga, "An Implementation of Handshake Join on FPGA", Proc. Second International Conference on Networking and Computing (ICNC 2011), Osaka, Japan, pp. 95-104, December 2011.
- C8. Takahiro KOMAMIZU, Toshiyuki AMAGASA, and Hiroyuki KITAGAWA, "A Framework of Faceted Navigation for XML Data", Proc. 13th International Conference on Information Integration and Web-based Applications & Services (iiWAS2011), Ho Chi Minh City, Vietnam, pp. 28-35, December 5-7, 2011.
- C9. Jianquan Liu, Hanxiong Chen, Kazutaka Furuse, Hiroyuki Kitagawa, and Jeffrey Xu Yu, "On Efficient Distance-based Similarity Search", Proc. 11th IEEE International Conference on Data Mining (ICDM), PhD Forum, Vancouver, Canada, pp. 1199-1202, December 11-14, 2011.
- C10. Mariko Kamie, Takako Hashimoto, and Hiroyuki Kitagawa, "Effective Web Video Clustering using Playlist Information", Proc. 2012 ACM Symposium on Applied Computing (SAC). (to appear)
- C11. Taiga Abe, Hideyuki Kawashima and Hiroyuki Kitagawa, "An Efficient Stream Archiving Method by Operator Merge and Write Control", Proc. 5th. International Workshop on Data Management for Wireless and Pervasive Communications. (to appear)

(3) 学会発表

- P1. 小山田昌史, 川島英之, 北川博之, "ストリームに起因する連続的トランザクション呼び出しの効率化", 信学技報, vol. 111, no. 76, DE2011-2, pp. 7-12, 2011年6月.
- P2. 小澤佑介, 天笠俊之, 北川博之, "GPGPU を用いた不確定データベースからの高速な頻出アイテム集合マイニング", 信学技報, vol. 111, no. 76, DE2011-9, pp. 49-54, 2011年6月.
- P3. 劉健全, 陳漢雄, 北川博之, "特化した R 木空間索引を用いた効率的な逆最遠傍検索", 第 4 回データ工学と情報マネジメントに関するフォーラム (DEIM 2012), D8-4, 2012年3月3日~3月5日.
- P4. 駒水 孝裕, 天笠 俊之, 北川 博之, "グラフデータに対するファセット探索のための頻出パターンを利用したオブジェクト抽出手法", 第 4 回データ工学と情報マネジメントに関するフォーラム (DEIM 2012), D3-3, 2012年3

月 3 日～3 月 5 日.

- P5. Salman Ahmed Shaikh, Hiroyuki Kitagawa, "Outlier Detection on Uncertain Data of Gaussian Distribution", 第 4 回データ工学と情報マネジメントに関するフォーラム(DEIM 2012), D3-1, 2012 年 3 月 3 日～3 月 5 日.
- P6. 大澤昇平, 天笠俊之, 北川博之, "RDFS Induction: URI の共通性を利用した Linked Data に対するスキーマ推定手法", 第 4 回データ工学と情報マネジメントに関するフォーラム(DEIM 2012), E11-6, 2012 年 3 月 3 日～3 月 5 日.
- P7. 上江まり子, 橋本隆子, 北川博之, "再生リストを利用した Web 動画クラスタリング手法の提案", 第 4 回データ工学と情報マネジメントに関するフォーラム(DEIM 2012), F6-1, 2012 年 3 月 3 日～3 月 5 日.
- P8. 張帆, 天笠俊之, 北川博之, 佐久間淳, "プライバシー保護オントロジーマッピングの提案", 第 4 回データ工学と情報マネジメントに関するフォーラム(DEIM 2012), B7-6, 2012 年 3 月 3 日～3 月 5 日.
- P9. 王春永, 川島英之, 北川博之, "衛星データからのホットスポットの抽出", 第 4 回データ工学と情報マネジメントに関するフォーラム(DEIM 2012), A7-4, 2012 年 3 月 3 日～3 月 5 日.
- P10. 清野真奈, 天笠俊之, 北川博之, "XML ストリームに対する省電力を考慮した問合せ処理", 第 4 回データ工学と情報マネジメントに関するフォーラム(DEIM 2012), E11-4, 2012 年 3 月 3 日～3 月 5 日.
- P11. 山口祐人, 天笠俊之, 北川博之, "マイクロブログにおけるユーザ専門性を考慮した Web ページ推薦", 第 4 回データ工学と情報マネジメントに関するフォーラム(DEIM 2012), F3-5, 2012 年 3 月 3 日～3 月 5 日.
- P12. 小山田昌史, 川島英之, 北川博之, "ストリームデータ処理における状態一貫性の保証", 第 4 回データ工学と情報マネジメントに関するフォーラム(DEIM 2012), B8-2, 2012 年 3 月 3 日～3 月 5 日.
- P13. 小澤佑介, 天笠俊之, 北川博之, "GPU による不確実データベースからの確率的頻出アイテム集合マイニングの高速化", 第 4 回データ工学と情報マネジメントに関するフォーラム(DEIM 2012), F6-5, 2012 年 3 月 3 日～3 月 5 日.
- P14. 高橋毅, 天笠俊之, 北川博之, "レビューデータにおける評価の時系列的変化に着目したイベント検出", 第 4 回データ工学と情報マネジメントに関するフォーラム(DEIM 2012), F6-2, 2012 年 3 月 3 日～3 月 5 日.
- P15. 富山克裕, 川島英之, 北川博之, "安全性を考慮したストリームデータ処理", 第 4 回データ工学と情報マネジメントに関するフォーラム(DEIM 2012), B8-5, 2012 年 3 月 3 日～3 月 5 日.
- P16. Chongjie LI, Toshiyuki AMAGASA, Hiroyuki KITAGAWA, "Label-bag based Graph Anonymization by Edge Addition", 第 4 回データ工学と情報

マネジメントに関するフォーラム(DEIM 2012), B7-3, 2012 年 3 月 3 日～3 月 5 日.

- P17. 井上寛之, 天笠俊之, 北川博之, "OLAP を利用した Linked Data の分析処理", 情報処理学会 第 74 回全国大会, 2N-2, 2012 年 3 月 6 日～3 月 8 日.
- P18. 坂倉悠太, 天笠俊之, 北川博之, "ソーシャルブックマークにおける複数のユーザアカウントを用いた集中的ブックマークの検出", 情報処理学会 第 74 回全国大会, 3N-4, 2012 年 3 月 6 日～3 月 8 日.
- P19. 中村高士, 早瀬康裕, 北川博之, "プロジェクト横断的なオープンソースソフトウェア開発記録の分析手法", 情報処理学会 第 74 回全国大会, 3L-9, 2012 年 3 月 6 日～3 月 8 日.
- P20. 西村和也, 川島英之, 北川博之, "メディアストリームとタプルストリームの統合的管理に関する研究", 情報処理学会 第 74 回全国大会, 1N-7, 2012 年 3 月 6 日～3 月 8 日.
- P21. 林史尊, 天笠俊之, 北川博之, 海老沢研, 中平聡志, "動的タイムワーピング距離を用いた X 線天文データの類似検索", 情報処理学会 第 74 回全国大会, 4N-8, 2012 年 3 月 6 日～3 月 8 日.
- P22. 栗原 耕平, 川島 英之, 北川 博之, "分散ストリーム処理環境における高信頼化手法の適用とその評価", 情報処理学会 第 74 回全国大会, 1N-4, 2012 年 3 月 6 日～3 月 8 日.

(4) 受賞

- A1. 学生奨励賞, 清野真奈, 天笠俊之, 北川博之, "XML ストリームに対する省電力を考慮した問合せ処理", 第 4 回データ工学と情報マネジメントに関するフォーラム(DEIM 2012), E11-4, 2012 年 3 月 3 日～3 月 5 日.

(5) その他成果デモ展示

- E1. イノベーションジャパン 2011 (主催: JST, NEDO): 「リアルタイムデータ分析を支援するストリームクラウド基盤」