VII. 計算情報学研究部門

VII-1. データ基盤分野

1. メンバー

教授　　北川　博之
准教授　　天笠　俊之
学生　　大学院生　32 名、学類生　7 名

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一的に扱うための 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，産業技術総合研究所，JAXA 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

【１】　情報統合基盤技術
（関連研究費：文部科学省受託研究，大川情報通信基金，三菱電機受託研究）
（１）指定イベント駆動型ストリーム処理

近年，ネットワークパケットやログデータやセンサデータ等，永続的に生成されるストリームデータが増加してきている．このストリームデータを扱うために，ストリーム処理エンジンが開発されている．従来のストリーム処理エンジンは，どのストリームソースから来たデータに対しても問合せが実行され，結果が生成される．しかし，応用処理によってはこれは必ずしも望ましいと言えないこともある．例えば，複数のセンサストリームを統合処理している場合，あるセンサ値により異常が観測された場合に他の関連センサデータを集約分析したいというような場合，全てのセンサ値に連動して統合処理を行うことは無駄である．また，統合処理においてはタイマー等に連動して周期的にデータの分析や集約を行いたいという場合も多い．このような要求に対応するためには，従来のようなあらゆる種類の新規データの到着に処理が連動するのではなく，特定の指定イベントのみに連動した指定イベント駆動型のストリーム処理が必要である．また，指定イベント駆動型のストリーム処理は，ストリーム型データ分析処理とバッチ型データ分析処理を自然に融合する上でも重要である．

本研究では，指定イベント駆動型ストリーム処理を実現したストリーム統合基盤システム JsSpinner のプロトタイプを開発した．JsSpinner は，多様な不均質データの統合処理に対応するため，多くのストリーム処理エンジンが採用しているリレーショナルストリームではなく，半構造データの JSON 形式のストリームや情報源を対象とするように設計されている．指定イベント駆動型ストリーム処理では，問合せにおいてユーザによってマスタストリームとして指定されたストリームソースから新規のデータが到着したときのみ，問合せが評価される．

実際のストリーム処理エンジンにおける問合せ処理では，冗長な処理を避けるため差分計算が行われる．指定イベント駆動型ストリーム処理を差分計算処理の枠組みの中で実現する最もナイーブな方法は，マスターストリームからの新規タプルに起因する結果か，それ以外の新規タプルに起因する結果かにマーキングを行い，マスタマークのあるタプルが到着した時点でのみ，最終的な問合せ結果の出力を行うというものである．しかし，ナイーブな方式では，最終的な問合せ結果に寄与しない多くの無駄な処理が発生し得る点が問題としてある．そこで本研究では，マスタストリームデータを処理するためのマスタウィンドウ演算子と非マスタストリームデータを処理するためのスマートウィンドウ演算子を導入するスマート方式を提案した．予備的な比較実験の結果はスマート方式の有効性を確認した．

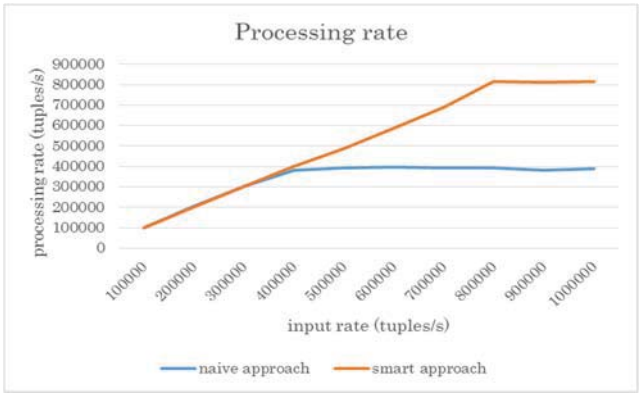


図 1　スマート方式とナイーブ方式の比較

（2）トランザクショナルストリーム処理

データストリームの分析処理のためにデータストリーム処理エンジンが用いられている．データストリームのより詳細な解析処理のためには，データストリームであるログデータとリレーショナルデータベースなどの外部リソース中のユーザデータを結合する処理などが必要になる場合がある．しかしこのような処理を実行する際に外部リソースに対して更新処理が施される場合を考えると，連続的問合せの一回の処理結果の中で，一貫して同じ外部リソースの状態を参照しなくなるという一貫性の問題が発生しうる．そこで，我々は連続的問合せの一回の処理結果における外部リソースの参照が一貫性を有していることを保証する，トランザクショナルストリーム処理の概念を既に提案している．本研究では，差分計算処理の枠組みにおいてトランザクショナルストリーム処理を実現するための方式を提案し，その有効性を検証した．具体的には，外部リソースの更新を監視するモニタ演算子を導入し，更新イベントをストリームとして下流に伝搬することで参照の一貫性を保

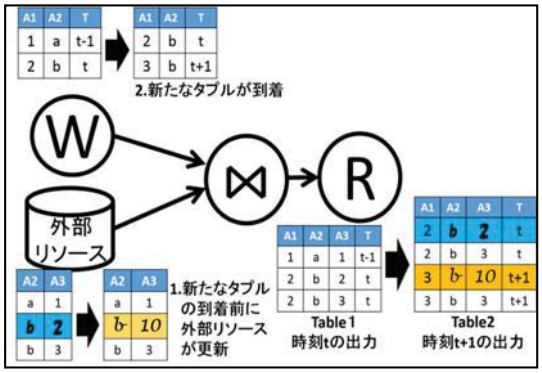


図 2　一貫性のない外部リソース参照

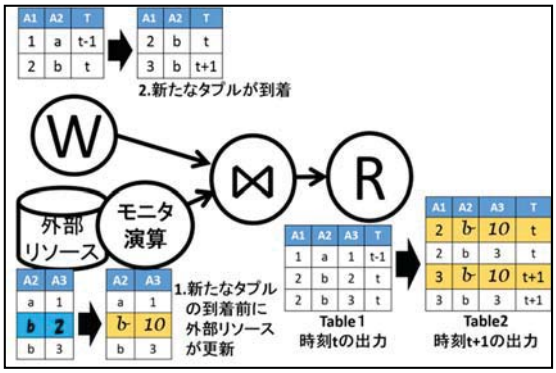


図 3　一貫性のある外部リソース参照

（3）ストリーム OLAP

近年，センサデータやマイクロブログ等，連続的に生成され，配信されるようなストリームデータの増加に伴い，ストリーム処理エンジンが多数開発されてきている．一方，ストリームデータに対してより高いレベルの分析を行いたいというニーズが増加している．この代表例として，多次元データ分析(OLAP 処理)がある．ストリームデータに対して

OLAP 処理を行う研究として, Jiawei Han らはストリームデータに対する OLAP 処理を容易にする Stream Cube を提案している. ただし, ストリーム処理エンジンを活かした OLAP 処理については十分検討が行われていない.

そこで本研究では，ストリーム処理エンジンを用いた OLAP 処理を実現するための手法を検討した．具体的には，エンジン側に連続的問合せを登録しておき，これらの登録しておいた問合せから OLAP 処理の結果として必要なデータを得る．OLAP 処理では，分析対象のデータの次元（属性）や次元内の階層の全ての組合せを頂点とする lattice を考える．例えば，商品情報（商品 ID, 商品名，ジャンル，顧客 ID, 顧客名，地域, 売上額）のようなスキーマを考えると図のような lattice を構築できる．

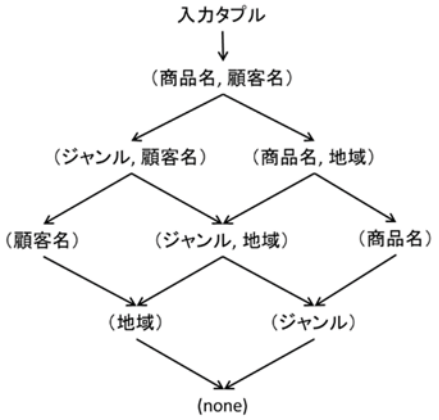


図 4　多段の集約レベル

lattice の各頂点は，OLAP 処理において集約する対象の次元の組合せに対応する．最も単純には，lattice 内の全ての頂点の次元の組合せの集約を実行すればよい．しかし一般には，lattice の全頂点数は膨大な数になることが多く，これらの全てをエンジンに登録することは，処理性能上難しい場合が多い．また，必ずしも全ての頂点に対応する集約値を常時取得し続けなくても，ユーザ要求に応じて導出できればよいという場合も多いと考えられる．そこで本研究では，lattice 内の一部の頂点のみを適切に選択し連続的問合せとして登録し，それ以外はユーザから求められたときのみ集約処理をし結果を生成する問合せ（オンデマンドに評価する問合せ）とすることで，効率的な OLAP 処理を実現する．

どの問合せをエンジンに登録し，どの問合せをオンデマンドに評価する問合せとするかの最適な組み合わせの選択は一般に組合せ最適化問題となるため，貪欲アルゴリズムを提案した．このアルゴリズムは，最初に全ての集約問合せをオンデマンドに評価する問合せとしておき，それぞれをエンジンに登録したときの処理コストと登録する前の処理コストの差分が最も大きくなるものを順に選んでいく．また，空間コストとして，エンジンに登録した集約問合せが保持する演算対象のタプル数と演算結果のタプル数の総数とする．

【２】　データマイニング・知識発見技術

（関連研究費 : 文部科学省受託研究, 産総研基盤研究(A), 富士通研究所受託研究）

（１）不確実データに対する外れ値検出

データや応用の多様化や，各種センサデバイスの発達に伴い，不確実性を伴うデータ処理に対する要求が高まっている．例えば，GPS による位置情報には，本質的に誤差が含まれており，その誤差を考慮した処理が求められる. 一方, 通常のデータからは著しく異なるデータ（外れ値）を検出する外れ値検出がさまざまな応用で利用されている．不確実データに対して外れ値検出を行う場合，データの不確実性を考慮した上で検出処理を行うことが望ましい.

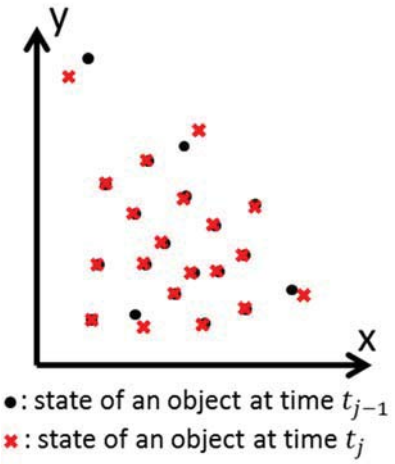


図 5　ストリームにおける外れ値

本研究では, ガウス分布に従う不確実性を持つデータに対する距離に基づく外れ値検出手法について検討した. 特に, 外れ値の度合いが大きいものから k 件の外れ値を検出するトップ k 外れ値検出のアルゴリズムを考案しその有効性を示した．厳密な外れ値度を計算する上ではガウス分布を考慮した距離計算が必要であるが, それには多大な計

算コストが伴う．そこで，提案手法では，外れ値候補オブジェクトの外れ値度の上限と下限を求め，それに基づきトップ k 外れ値に入り得るかを判定する k とで，計算コストを大幅に削減する．

またさらに，静的なデータ集合ではなく，データがストリームとして系列的に到着する場合を想定し，それに対応した差分計算に基づく効率的な連続的外れ値検出の手法についても検討を行った．

（2）Twitter ユーザの位置推定

近年では，ソーシャルメディアの普及により，ソーシャルストリームの情報を利活用して実世界をモニタリングする研究が多く行われている．ソーシャルストリームの情報を利用する上で，時間と位置に関する情報が重要である．時間情報は比較的容易に得ることができるのに対し，位置情報の把握はかならずしも容易ではない．代表的な位置情報として情報発信者の居住地情報がある．しかし，多くの研究で指摘されているように，ソーシャルメディアユーザは自らの居住地を公開していないことが多い．本研究では，ストリーム情報源に対するメタデータ推定の具体的な事例として，マイクロブログユーザの居住地推定手法を開発した．

マイクロブログユーザの居住地推定を行う従来の研究は，大きくグラフベース手法とコンテンツベース手法に分類される．前者はユーザの友人関係等を示したグラフの分析がベースとなっており，後者は発信されたコンテンツの分析がベースとなっている．本研究においては，グラフベースとコンテンツベースの両手法に関して従来よりもより推定精度の高い手法を開発した．

○ソーシャルグラフにおけるグラフランドマークを用いた手法

従来のグラフベース手法の大部分は closeness assumption を基にしてユーザの居住地を推定している．Closeness assumption とは，ソーシャルグラフ上で接続されているユーザ同士（友人等）はその居住地が互いに近いという仮定である．しかし，ソーシャルグラフの性質によっては closeness assumption は必ずしも有効ではない．例えばある一般ユーザのアカウントと，そのユーザが興味を持っている企業などのアカウントがソーシャルグラフ上で接続されることがある．このような環境では closeness assumption はあまり有効でないことが多い．

そこで本研究では closeness assumption とは異なる concentration assumption を導入する．Concentration assumption とは，ソーシャルグラフ上には自らのフォロワー群の居住地がある地域に集中しているユーザが存在するという仮定である．このユーザのことをグラフランドマークと呼ぶ．グラフランドマークを用いると，「グラフランドマークのフォロワー群の居住地は互いに近い」という推定が可能になる．

提案手法（landmark mixture model; LMM）は，グラフランドマークをフォローするユーザ群の居住地は互いに近いという仮定にもとづき，ユーザの居住地を確率分布でモデル化する手法である．まず全てのユーザに対して，そのユーザのフォロワー群の居住地の分布（dominance distribution）を計算し，割り当てる．そして，あるユーザの居住地の分布をそのユーザがフォローするユーザ群の dominance distribution を混合することにより得る．得られた居住地の分布において確率密度が最大になる点を推定した居住地とする．

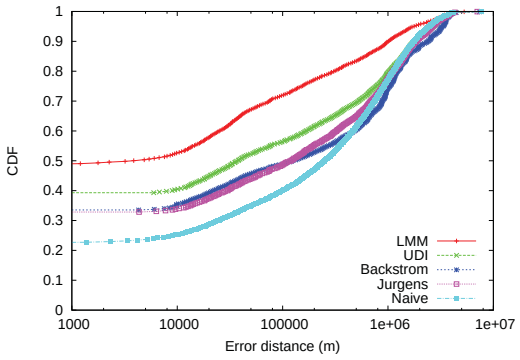


図 6　比較実験結果

比較実験の結果を図 6 に示す．横軸は推定誤差（単位:m）の値を示し，縦軸は推定誤差が対応する値以下であるユーザの割合，すなわち精度を示している．提案手法が他の手法を概ね上回っていることが分かる．

○ソーシャルストリームを用いたオンライン居住地推定

ソーシャルストリームでは，ユーザが時々刻々と大量の投稿をしている．そのため，居住地推定の手がかりとなる情報もリアルタイムに増え続けており，コンテンツベース手法においても本来インクリメンタルに居住地を推定することが可能である．しかし，既存のコンテンツベース手法では，このようなインクリメンタルな処理は実現できていない．

本研究では，ソーシャルストリームから次々に得られるコンテンツを基に居住地を逐次推定することの出来る手法（Online Location Inference Method; OLIM）を提案した．また，本手法ではコンテンツの時間的特徴を考慮することにより，temporally-local word という新しいローカルワードを導入する．Temporally-local word とは，従来の定常的な局所性を持つローカルワード（statically-local word）とは異なり，一時的な局所性を持つ単語のことである．本手法ではこれら 2 種類のローカルワードを併用する．

比較のための評価実験では，提案手法と四つの既存手法（UDI, Cheng, Hecht, Kinsella）及びナイーブな手法（NaiveC）を比較した．ナイーブな手法とは，ユーザが投稿した地名のメドイドを計算する手法である．また，提案手法は二つのローカルワードを併用するものと，statically-localwords のみを用いるものとを比較した．

図 7 は各手法の精度の比較結果を示している．結果から，提案手法の精度が最も高いことが分かる．図 8 は時間を追って提案手法が推定誤差を減少させていく結果を示している．横軸はデータセットに含まれるツイートを時系列順に処理した時の処理したツイートの割合を表し，縦軸はその辞典での推定誤差の中央値を表している．結果から，推定誤差はオンライン推定により徐々に減少していくことが分かる．

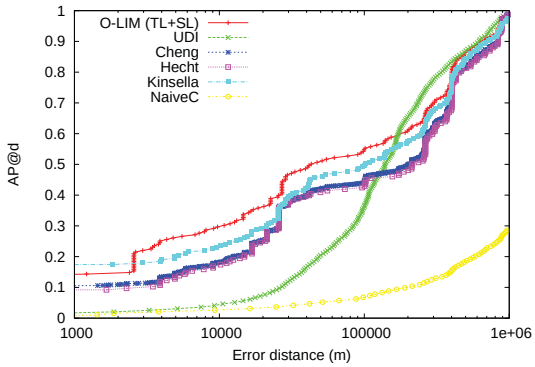


図 7　推定精度の比較

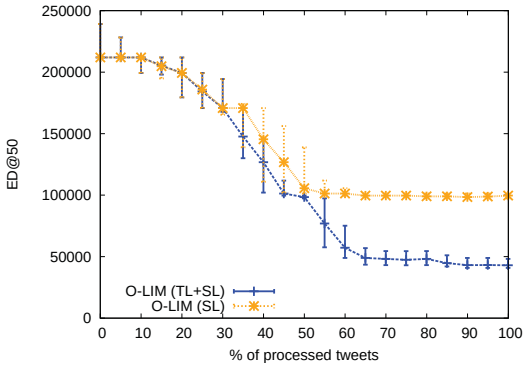


図 8　推定精度の時間変化

（3）GPU を用いた不確実トランザクションデータに対する確率的頻出アイテム集合マイニング

不確実性を含む大量のデータの処理のために，不確実データベースの研究が広く行なわれている．不確実データベースに対して，頻出アイテム集合マイニングを行なう手法がいくつか提案されているが，処理速度が遅いという問題がある．一方，GPU (Graphics Processing Unit) を用いた GPGPU (General Purpose computation on GPU)という手法が，高性能計算の分野で注目されている．GPGPU は，元々はグラフィック処理のための演算装置である

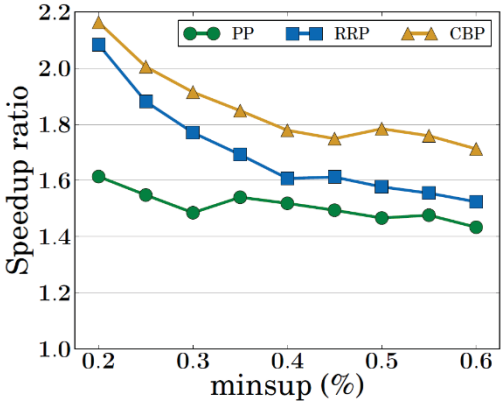


図 9 性能評価 (Kosarak データセット)

GPU を，その高い並列度をいかして汎用的な計算に利用するものである．

我々はこれまで単一 GPU を用いた不確実データベースに対する頻出アイテム集合マイニングの高速化のための手法を提案してきた．本研究では単一 GPU に対する手法をベースに，複数の GPU を搭載した複数のノードから構成されるクラスタ環境において，同様の処理を行う手法を提案する．複数 GPU を用いる利点として，GPU の増加による並列度の向上がある．また，GPU のメモリはそれほど大きくないが，複数 GPU にすることで利用可能なメモリ領域が増加し，巨大なデータを扱うことも可能となる．一方，複数 GPU を利用する上で問題になる点として，GPU 間でのデータ通信がある．通常，GPU は PCI-Express によって接続されているが，ここの通信のメモリバンド幅が小さいため，ボトルネックになる可能性が高い．そのため，GPU 間でのデータの依存をできるだけ削減することが望ましい．提案手法では，GPU 間でのデータ通信の削減に加え，負荷分散を行ない処理の高速化を図った．さらに，複数 GPU を持つノードからなるクラスタにおける手法も提案した．また，実験により，提案手法の性能を評価した．図 9 が実験結果である．NVIDIA Tesla M2090 を 2 台搭載した単一ノードの実験において，単一 GPU を利用した場合に比べて最大 2 倍程度の高速化が実現できている．また，スペースの都合で割愛するが，複数ノードを利用した環境でも高速化が達成されることを示した．

【３】 XML・Web プログラミング

（１） LINQ を用いた LOD 問合せ

政府や行政などの公共機関が保有する様々なデータを，再利用可能な形式で外部へ公開する取り組みをオープンデータと呼び，近年先進国を中心に積極的に推進されている．このような構造化されたデータを Web 上で公開，共有する一つの方法として Linked Open Data（LOD）が注目されている．

LOD では，一般に RDF（Resource Description Framework）フォーマットが用いられ，RDF データに対する問合せには，専用の問合せ言語である SPARQL が用いられる．しかしながら SPARQL の記述においては，問合せ言語の習得，LOD および RDF の関連技術についての知識が必要になる．また，RDF は本質的にはグラフ構造であり，LOD は複雑かつ冗長な構造になるため，これらの知識を持たない利用者が，LOD として公開されたオープンデータを利用するのは容易ではない．

そこで，本研究ではビューを用いた LINQ による LOD に対する問合せを提案した．予め LOD や RDF に知識のあるデータベース設計者らが LOD に対する JSON ビューを定義する．定義されたビューに対して，米マイクロソフト社が提供する.NET Framework の機能の一つである LINQ を用いて，C#等のプログラミング言語から LOD に対する問合せを記述する．より具体的には，まず設計者が，本手法で提案するビュー定義言語を用いて，ビュー定義を行う（図 10）．データの利用者は，ビュー定義に基づいて，LINQ 問合せを記述すると，システムが LINQ 問合せをビュー定義に基づき対応する情報源に対する SPARQL 問合せに書き換える．SPARQL 問合せの結果は，JSON 形式のデータに変換され，利用者に返却される．このように，利用者は SPARQL や LOD の詳細を知ることなく，LOD を利用することが可能となる．

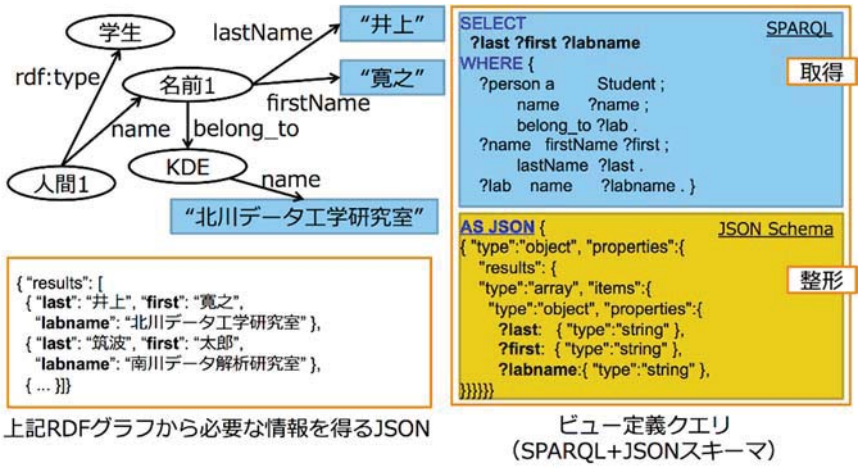


図 10 LOD に対する JSON ビューの定義

（2）局所計算に基づく ObjectRank 推定

PageRank はリンク構造解析手法の一つであり，グラフ上のノードの重要度を評価する手法である．Web 検索やソーシャルネットワーク分析，バイオインフォマティクスなどの様々な分野に用いられている．しかし，PageRank は計算コスト が大きいという問題がある．また，多くの応用において，全てのノードにおける PageRank スコアは必ずしも必要でなく，少数のノードのスコアだけ計算できれば十分であるケースが多い．

この問題に対して，グラフ全体の情報を用いずに特定のノード（対象ノード の PageRank 値を計算する手法が提案されている．本研究では，Chen らの手法をベースに，より効率的な手法を提案した．Chen の手法では，PageRank スコアを計算したい対象ノードに対して，その周囲の影響力の強い部分グラフを同定する．あるノードの影響力はそのノードがエッジを張っているノードの影響力を用いて再帰的に計算することができるという特徴を用いて，反復計算を行わずに部分グラフを同定することができる．これにより，提案手法は PageRank 値の推定精度を維持したまま，効率的に対象ノードの PageRank 値 を推定することができる．

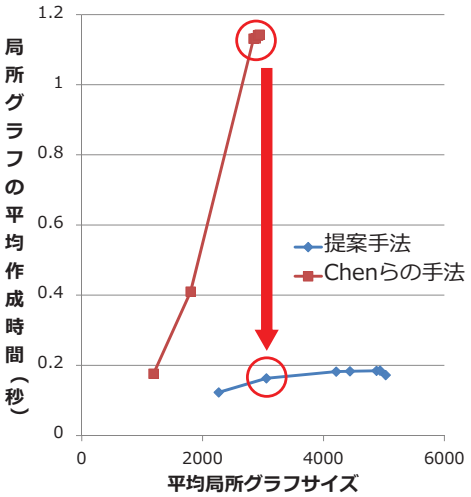


図 11 Web グラフを利用した実験結果

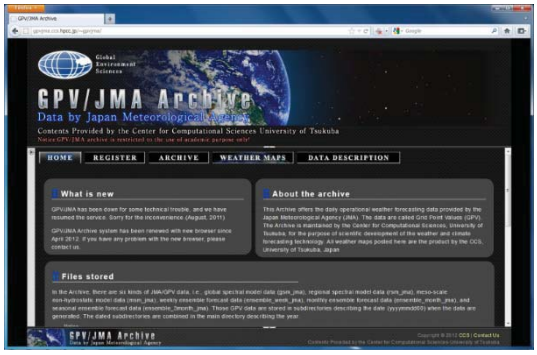
提案手法の性能を評価するために Twitter と Web グラフのデータセットを用いて評価実験を行った．実験結果より，提案手法は推定精度を維持しながら，高速に対象ノードの PageRank 値を推定できることが分かった．具体的には，Chen らの手法と比較して，提案手法は精度を維持したまま高速に PageRank 値を推定することができた（図 11）．

【4】 科学分野におけるデータベース応用

（1） GPV/JMA アーカイブ

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，季間アンサンブルの 6 種類である．



（2） 格子 QCD データグリッド
ILDG/JLDG

Japan Lattice Data Grid (JLDG), International Lattice Data Grid (ILDG)は，格子 QCD 配位データを共有するためのデータグリッドである. 素粒子物理研究部門と連解し, JLDG/ILDG の運営に継続参画している．

（3）X 線天体観測データにおけるアウトバーストの類似検索

ブラックホール，中性子星などは X 線を発する天体として知られており，それらの天体には，短期間に大量の X 線を放出する「アウトバースト」という現象が存在することが知られている．また，JAXA 宇宙研海老沢教授らのグループにより，異なる X 線天体の間で，アウトバーストの X 線強度変化に類似性が見られることが近年明らかにされた．これは，背後にある物理過程の類似性を示す可能性があり興味深い．このため，海老沢教授らのグループと共同で，X 線天体の観測データを対象に，類似したアウトバーストパターンを検索する手法を研究開発している．

4. 教育

学生の指導状況

【学位論文】

＜博士論文＞

1. Salman Ahmed SHAIKH
   A Study on Distance-based Outlier Detection on Uncertain Data
2. 村上 直
   概念モデリングに基づく O/R マッピング手法に関する研究
3. 山口 祐人
   A Study on User Location Inference in Social Media
4. 高橋 翼
   系列データの匿名化に関する研究
5. 丸橋 弘治
   A Study on Large Scale Graph Analysis Using Eigen Decomposition and Tensor Decomposition

＜修士論文＞

1. 西村 直孝

一括評価による複合イベント処理の高スループット化

2. 坂倉 悠太
   リンク構造解析における部分グラフに基づいた効率的なノード評価値の推定
3. 井上 寛之
   Linked Open Data に対する多様な問合せ処理に関する研究
4. 中村 高士
   リポジトリを跨いだコミットトランザクションの推定に基づくロジカルカップリング検出手法
5. Bou Savong
   A Study on Keyword Search over XML Streams
6. 石 剣峰
   Modularity-based Clustering of Dynamic Graphs
7. 林 史尊
   GPU を用いた Canopy クラスタリングの高速化

＜特定課題研究報告書＞

財前 涼
字幕ダウンロード機能付き HDD レコーダ開発による字幕後付システムの実現
-HDD レコーダの GUI 及びサーバとの通信機能の実装-

＜学士論文＞

1. 岡田 莉奈
   ソーシャルネットワークデータの距離関係の変化を抑制する k 匿名化アルゴリズム
2. 森 智彦
   ソーシャルリーディングシステムにおけるデータの格納と検索に関する研究
3. 大西 誠
   Twitter の即時性に着目したニュース記事のリアルタイムソーシャルアノテーション
4. 熊本 和正
   テンソル分解によるユーザレビューの分析に関する研究
5. 小柳 涼介
   XML データにおけるテキストおよび構造を考慮した効率的な類似検索

5. 受賞、外部資金、知的財産権等

＜受賞＞

A1. 学生プレゼンテーション賞, 坂倉 悠太, 山口 祐人, 天笠 俊之, 北川 博之, "部分グラ

フに基づく効率的な PageRank 推定", 第 6 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 2014), D6-4, 2014 年 3 月 3 日～3 月 5 日.

A2. 学生プレゼンテーション賞, 林 史尊, 小澤 佑介, 天笠 俊之, 北川 博之, "GPU を用いた Canopy クラスタリングの高速化", 第 6 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 2014), D5-4, 2014 年 3 月 3 日～3 月 5 日.

A3. 学生プレゼンテーション賞, 優秀インタラクティブ賞, 岡田 莉奈, 渡辺 知恵美, 北川博之, "ソーシャルネットワークデータの距離関係の変化を抑制する k 匿名化アルゴリズム", 第 6 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 2014), E5-4, 2014 年 3 月 3 日～3 月 5 日.

A4. 学生プレゼンテーション賞, 小柳 涼介, 天笠 俊之, 北川 博之, "テキストおよび構造の類似度に基づいた XML データに対する効率的な類似検索", 第 6 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 2014), D7-5, 2014 年 3 月 3 日～3 月 5 日.

A5. 学生奨励賞, 大西 誠, 北川 博之, "ニュース記事の効率的なリアルタイムソーシャルアノテーション手法", 情報処理学会第 76 回全国大会 (IPSJ 全国大会 2014), 3M-8, 2014 年 3 月 11 日～3 月 13 日.

A6. 学生奨励賞, 小柳涼介, 天笠俊之, 北川博之, "大規模 XML データにおける効率的な重複データ検出", 情報処理学会第 76 回全国大会 (IPSJ 全国大会 2014), 3N-8, 2014 年 3 月 11 日～3 月 13 日.

A7. 日本データベース学会論文賞, 福角駿, 森嶋厚行, 品川徳秀, 北川博之, "DB 抽象化とゲーム理論に基づくマイクロブログからの構造データ抽出 GWAP の開発", 日本データベース学会論文誌, Vol. 11, No. 1, pp. 19-24, 2012.

＜外部資金＞

受託経費：文部科学省　　（平成 25 年度）

研究課題：ビックデータ利活用のためのデータ連携技術に関するフィージビリティスタディ及び予備研究

研究代表者：北川　博之

配分金額：27,405,000 円（直接経費 21,080,770：間接経費 6,324,230）

研究種目：基盤研究(A)　（平成 24 年度～平成 26 年度）

研究課題：大規模・異種の時空間データ統合で生じる矛盾を許容するサイエンスクラウド基盤（研究代表者：小島功（産総研））

研究分担者：北川　博之

配分金額：1,300,000 円（直接経費 1,000,000：間接経費 300,000）

研究分担者：天笠　俊之

配分金額：2,210,000 円（直接経費 1,700,000：間接経費 510,000）

寄付金：大川情報通信基金　　（平成 25 年度）
研究課題：大規模実世界実時間情報基盤のための高度ストリーム処理
研究代表者：北川　博之
配分金額：1,000,000 円（直接経費 950,000：間接経費 50,000）

受託経費：三菱電機株式会社　　（平成 25 年度）
研究課題：時系列データベース・分析技術の研究開発
研究代表者：北川　博之・天笠　俊之
配分金額：500,000 円（直接経費 450,000：間接経費 50,000）

受託経費：株式会社富士通研究所　　（平成 25 年度）
研究課題：時系列データの分析基盤技術の研究
研究代表者：北川　博之・天笠　俊之
配分金額：2,000,000 円（直接経費 1,800,000：間接経費 200,000）

研究種目：基盤研究(C)　（平成 25 年度～平成 27 年度）
研究課題：ＥＰＵ3.0 を核とした知識集積型ソーシャルリーディング基盤に関する研究
研究代表者：天笠　俊之
配分金額：1,690,000 円（直接経費 1,300,000：間接経費 390,000）

研究種目：基盤研究(A)　（平成 25 年度）
研究課題：災害後の復旧・復興における共有情報管理
研究分担者：天笠　俊之
配分金額：1,300,000 円（直接経費 1,000,000：間接経費 300,000）

6. 研究業績

(1) 研究論文

A) 査読付き論文

＜学術雑誌論文＞

J1. Rong-Hua Li, Jianquan Liu, Jeffrey Xu Yu, Hanxiong Chen, and Hiroyuki Kitagawa, "Co-occurrence Prediction in a Large Location-based Social Network", Frontiers of Computer Science, Vol. 7, No. 2, pp. 185-194, April 2013.

J2. 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", 宇宙科学情報解析論文誌 第二号, pp. 19-27,

June 2013.

J3. Masafumi Oyamada, Hideyuki Kawashima, and Hiroyuki Kitagawa, "Data Stream Processing with Concurrency Control", SIGAPP Appl. Comput. Rev., Vol. 13, No. 2, pp. 54-65, June 2013.

J4. 山口祐人, 伊川洋平, 天笠俊之, 北川博之, "ソーシャルメディアにおけるローカルイベントを用いたユーザ位置推定手法", 情報処理学会論文誌データベース (TOD60), Vol. 6, No. 5, pp. 23-37, December 2013.

J5. 丹治寛佳, 森嶋厚行, 井ノ口宗成, 北川博之, "Web 情報を用いた竜巻経路推定支援のためのクラウドソーシング技術開発の試み", 情報処理学会論文誌データベース (TOD60), Vol. 6, No. 5, pp. 95-106, December 2013.

＜査読付き国際会議論文＞

C1. Yutaka Kabutoya, Tomoharu Iwata, Hiroyuki Toda, and Hiroyuki Kitagawa, "A Probabilistic Model for Diversifying Recommendation Lists", Proc. 15th Asia-Pacific Web Conference (APWeb 2013), Sydney, Australia, pp. 348-359, April 4-6, 2013.

C2. Salman Shaikh and Hiroyuki Kitagawa, "Fast Top-k Distance-based Outlier Detection on Uncertain Data", Proc. 14th International Conference on Web-Age Information Management (WAIM 2013), Beidaihe, China, LNCS7923, pp. 301-313, June 14-16, 2013.

C3. Hiroyuki Inoue, Toshiyuki Amagasa, and Hiroyuki Kitagawa, "An ETL Framework for Online Analytical Processing of Linked Open Data", Proc. 14th International Conference on Web-Age Information Management (WAIM 2013), Beidaihe, China, LNCS7923, pp. 111-117, June 14-16, 2013.

C4. Tadashi Murakami, Toshiyuki Amagasa, and Hiroyuki Kitagawa, "DBPowder: A Flexible Object-Relational Mapping Framework based on a Conceptual Model", Proc. 37th IEEE International Computer Software and Applications Conference (COMPSAC 2013), Kyoto, Japan, pp. 589-598, July 22-26, 2013.

C5. Yusuke Kozawa, Toshiyuki Amagasa, and Hiroyuki Kitagawa, "Parallel and Distributed Mining of Probabilistic Frequent Itemsets Using Multiple GPUs", Proc. 24th International Conference on Database and Expert Systems Applications (DEXA 2013), Prague, Czech Republic, pp. 145-152, August 26-30, 2013.

C6. Kenji Gonnokami, Atsuyuki Morishima, Shigeo Sugimoto, Hiroyuki Kitagawa, "Condition-Task-Store: A Declarative Abstraction for Microtask-based Complex Crowdsourcing", Proc. First VLDB Workshop on Databases and Crowdsourcing (DBCrowd 2013), pp. 20-25, August 26-30, 2013.

C7. Kousuke Nakabasami, Toshiyuki Amagasa, and Hiroyuki Kitagawa, "Querying MongoDB with LINQ in a Server-side JavaScript Environment", Proc. 2nd International Workshop on Advances in Data Engineering and Mobile Computing (DEMoC 2013), Gwangju, Korea, pp. 344-349, September 4-6, 2013.

C8. Yuto Yamaguchi, Toshiyuki Amagasa, and Hiroyuki Kitagawa, "Landmark-Based User Location Inference in Social Media", Proc. 1st ACM Conference on Online Social Networks (COSN 2013), Boston, USA, pp. 223-234, October 7-8, 2013.

C9. Atsuyuki Morishima, Erika Yumiya, Masami Takahashi, Shigeo Sugimoto, Hiroyuki Kitagawa, "Efficient Filtering and Ranking Schemes for Findin Inclusion Dependencies on the Web", Proc. 22nd International Conference on Information and Knowledge Management (CIKM), pp. 763-768, October 29, 2013.

C10. Yuta Sakakura, Yuto Yamaguchi, Toshiyuki Amagasa, and Hiroyuki Kitagawa, "A Local Method for ObjectRank Estimation", Proc. 15th International Conference on Information Integration and Web-based Applications & Services (iiWAS 2013), Vienna, Austria, pp. 92-101, December 2-4, 2013.

C11. Eri Kataoka, Toshiyuki Amagasa, and Hiroyuki Kitagawa, "A System for Social Reading based on EPUB3", Proc. 15th International Conference on Information Integration and Web-based Applications & Services (iiWAS 2013), Vienna, Austria, pp. 72-76, December 2-4, 2013.

(2) 国際会議発表

A) 招待講演

該当なし

B) 一般講演

C1. Yutaka Kabutoya, Tomoharu Iwata, Hiroyuki Toda, and Hiroyuki Kitagawa, "A Probabilistic Model for Diversifying Recommendation Lists", Proc. 15th Asia-Pacific Web Conference (APWeb 2013), Sydney, Australia, pp. 348-359, April 4-6, 2013.

C2. Salman Shaikh and Hiroyuki Kitagawa, "Fast Top-k Distance-based Outlier Detection on Uncertain Data", Proc. 14th International Conference on Web-Age Information Management (WAIM 2013), Beidaihe, China, LNCS7923, pp. 301-313, June 14-16, 2013.

C3. Hiroyuki Inoue, Toshiyuki Amagasa, and Hiroyuki Kitagawa, "An ETL Framework for Online Analytical Processing of Linked Open Data", Proc. 14th International Conference on Web-Age Information Management (WAIM 2013), Beidaihe, China, LNCS7923, pp. 111-117, June 14-16, 2013.

C4. Tadashi Murakami, Toshiyuki Amagasa, and Hiroyuki Kitagawa, "DBPowder: A Flexible Object-Relational Mapping Framework based on a Conceptual Model", Proc. 37th IEEE International Computer Software and Applications Conference (COMPSAC 2013), Kyoto, Japan, pp. 589-598, July 22-26, 2013.

C5. Yusuke Kozawa, Toshiyuki Amagasa, and Hiroyuki Kitagawa, "Parallel and Distributed Mining of Probabilistic Frequent Itemsets Using Multiple GPUs", Proc. 24th International Conference on Database and Expert Systems Applications (DEXA 2013), Prague, Czech Republic, pp. 145-152, August 26-30, 2013.

C6. Kenji Gonnokami, Atsuyuki Morishima, Shigeo Sugimoto, Hiroyuki Kitagawa, "Condition-Task-Store: A Declarative Abstraction for Microtask-based Complex Crowdsourcing", Proc. First VLDB Workshop on Databases and Crowdsourcing (DBCrowd 2013), pp. 20-25, August 26-30, 2013.

C7. Kousuke Nakabasami, Toshiyuki Amagasa, and Hiroyuki Kitagawa, "Querying MongoDB with LINQ in a Server-side JavaScript Environment", Proc. 2nd International Workshop on Advances in Data Engineering and Mobile Computing (DEMoC 2013), Gwangju, Korea, pp. 344-349, September 4-6, 2013.

C8. Yuto Yamaguchi, Toshiyuki Amagasa, and Hiroyuki Kitagawa, "Landmark-Based User Location Inference in Social Media", Proc. 1st ACM Conference on Online Social Networks (COSN 2013), Boston, USA, pp. 223-234, October 7-8, 2013.

C9. Atsuyuki Morishima, Erika Yumiya, Masami Takahashi, Shigeo Sugimoto, Hiroyuki Kitagawa, "Efficient Filtering and Ranking Schemes for Findin Inclusion Dependencies on the Web", Proc. 22nd International Conference on Information and Knowledge Management (CIKM), pp. 763-768, October 29, 2013.

C10. Yuta Sakakura, Yuto Yamaguchi, Toshiyuki Amagasa, and Hiroyuki Kitagawa, "A Local Method for ObjectRank Estimation", Proc. 15th International Conference on Information Integration and Web-based Applications & Services (iiWAS 2013), Vienna, Austria, pp. 92-101, December 2-4, 2013.

C11. Eri Kataoka, Toshiyuki Amagasa, and Hiroyuki Kitagawa, "A System for Social Reading based on EPUB3", Proc. 15th International Conference on Information Integration and Web-based Applications & Services (iiWAS 2013), Vienna, Austria, pp. 72-76, December 2-4, 2013.

(3) 国内学会・研究会発表

A) 招待講演

該当なし

B) その他の発表

P1. 福角駿, 森嶋厚行, 品川徳秀, 北川博之, "DB 抽象化とゲーム理論に基づくマイクロブログからの構造データ抽出 GWAP の開発", 日本データベース学会, 2013 年 6 月 22 日

P2. Salman Ahmed Shaikh, Hiroyuki Kitagawa, "Differential Outlier Detection on Uncertain Streams of the Gaussian Distribution", The 5th International Workshop with Mentors on Databases, Web and Information Management for Young Researchers (iDB2013), Sapporo, Japan, July 21 - 23, 2013.

P3. Yuto Yamaguchi, Toshiyuki Amagasa, Hiroyuki Kitagawa, "Landmark-Based User Location Inference on Social Media", The 5th International Workshop with Mentors on Databases, Web and Information Management for Young Researchers (iDB2013), Sapporo, Japan, July 21 - 23, 2013.

P4. Yuta Sakakura, Yuto Yamaguchi, Toshiyuki Amagasa, Hiroyuki Kitagawa, "A Local Method for ObjectRank Estimation", The 5th International Workshop with Mentors on Databases, Web and Information Management for Young Researchers (iDB2013), Sapporo, Japan, July 21 - 23, 2013.

P5. Hiroyuki Inoue, Toshiyuki Amagasa, Hiroyuki Kitagawa, "An ETL Framework for Online Analytical Processing of Linked Open Data", The 5th International Workshop with Mentors on Databases, Web and Information Management for Young Researchers (iDB2013), Sapporo, Japan, July 21 - 23, 2013.

P6. 中村 高士, 早瀬 康裕, 北川 博之, "ソフトウェアプロダクト間での Logical Coupling 検出に向けた予備的な調査", 第 182 回ソフトウェア工学研究発表会, 2013 年 10 月 24～10 月 25 日.

P7. 小澤 佑介, 天笠 俊之, 北川 博之, "データ分割と協調的マージに基づく GPU 上の効率的ソートアルゴリズム", 第 6 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 2014), A2-1, 2014 年 3 月 3 日～3 月 5 日.

P8. Savong Bou, Toshiyuki Amagasa, Hiroyuki Kitagawa, "Path-based Keyword Search over XML Streams", 第 6 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 2014), A1-5, 2014 年 3 月 3 日～3 月 5 日.

P9. 西村 直孝, 川島 英之, "リンク集約とパタンキャッシュを用いた複合イベント処理の高性能化", 第 6 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 2014), D3-3, 2014 年 3 月 3 日～3 月 5 日.

P10. 井上 寛之, 天笠 俊之, 北川 博之, "LINQ を用いた Linked Open Data に対する問合せ", 第 6 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 2014), D7-2, 2014 年 3 月 3 日～3 月 5 日.

P11. 坂倉 悠太, 山口 祐人, 天笠 俊之, 北川 博之, "部分グラフに基づく効率的な

PageRank 推定", 第 6 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 2014), D6-4, 2014 年 3 月 3 日～3 月 5 日.

P12. 林 史尊, 小澤 佑介, 天笠 俊之, 北川 博之, "GPU を用いた Canopy クラスタリングの高速化", 第 6 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 2014), D5-4, 2014 年 3 月 3 日～3 月 5 日.

P13. 王 岩, 北川 博之, "An Efficient Execution Scheme for Designated Event-based Stream Processing", 第 6 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 2014), D3-2, 2014 年 3 月 3 日～3 月 5 日.

P14. 黄 峻, 小澤 佑介, 天笠 俊之, 北川 博之, "GPGPU を用いた不確実時系列データ類似検索の高速化", 第 6 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 2014), D5-1, 2014 年 3 月 3 日～3 月 5 日.

P15. 中挾 晃介, 北川 博之, 天笠 俊之, "ストリーム処理エンジンを用いたストリームデータの OLAP 処理", 第 6 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 2014), D3-1, 2014 年 3 月 3 日～3 月 5 日.

P16. 岡田 莉奈, 渡辺 知恵美, 北川 博之, "ソーシャルネットワークデータの距離関係の変化を抑制する k 匿名化アルゴリズム", 第 6 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 2014), E5-4, 2014 年 3 月 3 日～3 月 5 日.

P17. 小柳 涼介, 天笠 俊之, 北川 博之, "テキストおよび構造の類似度に基づいた XML データに対する効率的な類似検索", 第 6 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 2014), D7-5, 2014 年 3 月 3 日～3 月 5 日.

P18. 権守健嗣, 森嶋厚行, 北川博之, "マイクロタスク型クラウドソーシング処理の変換", 第 6 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 2014), 2014 年 3 月 3 日.

P19. 福田 宏樹, 早瀬 康裕, 北川 博之, "クラス名の文法構造と周辺の識別子を用いたクラスの命名支援", 情報処理学会第 76 回全国大会 (IPSJ 全国大会 2014), 2M-5, 2014 年 3 月 11 日～3 月 13 日.

P20. 大西 誠, 北川 博之, "ニュース記事の効率的なリアルタイムソーシャルアノテーション手法", 情報処理学会第 76 回全国大会 (IPSJ 全国大会 2014), 3M-8, 2014 年 3 月 11 日～3 月 13 日.

P21. 小柳涼介, 天笠俊之, 北川博之, "大規模 XML データにおける効率的な重複データ検出", 情報処理学会第 76 回全国大会 (IPSJ 全国大会 2014), 3N-8, 2014 年 3 月 11 日～3 月 13 日.

P22. 岡田莉奈, 渡辺知恵美, 北川博之, "ノード間の距離関係を考慮したソーシャルネットワークにおける k 匿名化", 情報処理学会第 76 回全国大会 (IPSJ 全国大会 2014), 5N-5, 2014 年 3 月 11 日～3 月 13 日.

P23. 熊本和正, 天笠俊之, 丸橋弘治, 北川博之, "テンソル分解を用いたレビューデータの

分析", 情報処理学会第 76 回全国大会（IPSJ 全国大会 2014), 6N-7, 2014 年 3 月 11 日～3 月 13 日.

(4) 著書、解説記事等

井上克郎，楠本真二，後藤厚宏，鵜林尚靖，北川博之, "実践的情報教育協働ネットワーク enPiT", 情報処理, Vol.55, No.2, pp.194-197, February 2014.

7. 異分野間連携・国際連携・国際活動等

- 地球環境研究部門との連携：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，管理，運用.
- 素粒子物理研究部門との連携：Japan Lattice Data Grid (JLDG), International Lattice Data Grid (ILDG)の運営.
- 産業技術総合研究所との連携：大規模・異種の時空間データ統合で生じる矛盾を許容するサイエンスクラウド基盤に関する研究.
- Carnegie Mellon University との国際共同研究に関する準備.（2014 年度より共同研究実施中.）

8. 管理・運営

北川博之教授

- 学外
  - 文部科学省・情報技術人材育成のための実践教育ネットワーク形成事業「分野・地域を越えた実践的情報教育協働 NW」 ビジネスアプリケーション分野代表.
- 学内
  - システム情報工学研究科コンピュータサイエンス専攻：高度 IT・実践 NW 統括.
  - 計算科学研究センター：計算情報学研究部門長，計算科学振興室長.

天笠俊之准教授

- 学外
  - つくばライフサイエンス推進協議会 情報システム構築 WG 委員長
- 学内
  - CS 専攻・情報（科）学類学生委員会 委員長

9. 社会貢献・国際貢献

北川博之教授

- 国際委員等
  - 国際ジャーナル編集委員：IEEE Transactions on Knowledge and Data Engineering, World Wide Web Journal

- 国際会議運営委員：WAIM Steering Committee Member, DASFAA Steering Committee Member Emeritus
- 国際会議共同委員長：WAIM2013
- 国際会議共同最優秀論文委員長：DASFAA2013
- 国際会議共同パネル委員長：ASONAM2014
- 国際会議プログラム委員会委員：DASFAA2013, MDM2013, PAKDD2013, DEXA2013, IDEAS2013, CoopIS2013, DASFAA2014, MDM2014, PAKDD2014, IDEAS2014, DEXA2014
- 国際会議アドバイザリ委員：SRDS2014

- 国内委員等
  - 日本学術会議連携会員
  - 日本データベース学会副会長
  - （独）科学技術振興機構・戦略的創造研究推進事業「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化」領域アドバイザー
  - （独）情報通信研究機構・高度通信・放送研究開発委託研究評価委員会委員
  - 公益財団法人国際科学技術財団・国際科学技術財団 2014 年研究助成選考委員
  - FIT2014 第 13 回情報科学技術フォーラム現地実行副委員長

天笠俊之准教授

- 国際委員等
  - 国際学会プログラム委員：BSI2013, AICCSA 2013, WAIM2013, DEMoC2013, SITIS2013, FutureTech 2013, iiWAS2013
- 国内委員等
  - データ工学と情報マネジメントに関するフォーラム（DEIM フォーラム）2014 プログラム委員長
  - 電子情報通信学会論文誌「データ工学と情報マネジメント特集号」編集幹事
  - 日本データベース学会論文誌編集委員
  - 情報科学技術フォーラム（FIT）2013 プログラム委員

10. その他

特になし