

## 戦略的国際共同研究プログラム(SICORP)

日本－アメリカ共同研究

終了報告書 概要

- 研究課題名：「災害時応用のための効率的かつスケーラブルなビッグデータの収集・解析・処理」
- 研究期間：2015 年 4 月～2018 年 3 月
- 主な参加研究者名：  
日本側チーム

	氏名	役職	所属	研究分担
研究代表者	原 隆浩	教授	大阪大学	日本側担当の問題定義、方法論立案、機構の設計・拡張
研究参加者	西尾 章治郎	総長	大阪大学	日本側担当の機構の実装、実証実験
研究期間中の全参加研究者数			2名	

アメリカ側チーム

	氏名	役職	所属	研究分担
研究代表者	Sanjay Madria	Professor	Missouri University of Science and Technology	米国側担当の研究計画全般
研究期間中の全参加研究者数			1名	

#### 4. 共同研究の概要

本共同研究では、災害時において膨大なデータの収集・解析・処理を可能とする基盤技術確立することを目的として、(1) 高次元データの複数センサストリームの圧縮、(2) 高次元データ検索のための索引機構、(3) 複数の情報源から発生する高次元データのモニタリング、(4) マイクロログの実社会センシング応用という 4 つの課題に取り組んだ。

研究課題を推進する上で、課題(1)は、まず単独のセンサストリーム（ストリームは継続的に生成されるデータのこと）におけるクラウドを基盤とした協調圧縮機構を開発し、それをベースとして複数ストリームに拡張するため、2 つのワークパッケージ（WP1、WP4）として推進した。その他の課題はそれぞれ 1 つのワークパッケージ（WP2、WP3、WP5）とし、さらに考案技術・機構を実プラットフォーム上に実装し、評価実験を行うタスクを単独のワークパッケージ（WP6）とした。WP1、WP2、WP4 は主に米国チームが、WP3 と WP5 は日本チームがそれぞれ主導し、WP6 は両チームが密に連携を取りながら実施した。

- WP1：単独ストリームの協調データ圧縮機構
- WP2：高次元データの索引・検索機構
- WP3：複数の情報源からの高次元データのモニタリング機構
- WP4：複数ストリームの協調データ圧縮機構
- WP5：マイクロログの実社会センシング応用のための機構
- WP6：考案機構全体の実プラットフォーム上の実装・検証

#### 5. 共同研究の成果

##### 5-1 共同研究の学術成果

WP1 および WP4 では、ストリームデータを生成する複数の発生源（センサや IoT デバイス、クラウド）で協調的に圧縮を行う機構を開発した。WP2 では、大規模な高次元データに対して、複数の計算機の協調動作により、効率的に索引を構築し、代表的な検索技術

を実行する機構を開発した。WP3 では、複数の情報源から発生する大規模な高次元データに対して、効率的にモニタリングを行う機構について研究開発を実施した。考案技術は、従来技術と比較して、数 10～数 100 倍の高速性を実現した。WP5 では、マイクロログデータを実社会を表すセンサとして活用するために、マイクロログデータから利用価値の高い様々な知識を抽出する「社会センシング技術」と、異なる社会センシング技術に共通に利用できる「社会センシングのための基盤技術」の開発を行った。WP6 では、2 年目までに開発した技術・機構を統合したプラットフォームの実現を目指して研究開発を進めた。

本研究課題の WP1～WP5 では、互いの独立性を保ちつつ、頻繁に情報交換と議論を行うことで高いレベルでの技術開発を推進した。WP6 については、双方で開発したプラットフォームの完全な連携には至っていないが、両チームで連携して開発を継続している。

本研究課題は、学術的に大きな成果を達成した。研究期間全体で、学術論文 17 編、国際会議発表 26 件、招待講演 11 件、国内発表 32 件を公表した。これらの中には、国際的に最高権威もしくはそれに準ずる論文誌（ACM TOIS、IEEE TKDE（2 編）、ACM TSAS、WWW など）および国際会議（ICDCS、EMNLP、EDBT、Big Data、SSDM など）が多数含まれている。さらに、大阪科学賞などの権威のある研究賞を含め、6 件の賞を受賞した。

## 5-2 国際連携による相乗効果

両国の PI は、本研究課題を構成する要素技術について相補的かつ高い専門性を有しており、また、申請時点で既に 10 年以上の共同研究を行っている。このような状況で開始した本研究課題は、期待通りに、研究開始当初から二国間連携が円滑に行われた。その結果、幅広い分野をカバーすることが要求される技術・機構群を高いレベルで構築できた。成果の一部として、複数の共著論文が高いレベルの国際会議に採択され、また単独チームの論文は国際的に最高権威の論文誌や国際会議に多数採択されるに至った。これらの業績は、本共同研究の相乗効果の大きさを如実に表している。

さらに、実システムの開発経験が豊富な両者が連携することにより、センサ・IoT 系データと SNS データの両者をビッグデータ解析に利用可能なプラットフォームを構築できた。

本研究の初期段階には、両 PI により NII 湘南会議を主催し、4 か国から計算機科学、工学、地理学、人流解析学などの異なる分野の研究者・技術者 21 名が一堂に会して、災害とビッグデータに関する重要課題を議論した。さらに本研究の最終段階では、アウトリーチ活動もかねて米国 PI が主催した国際ワークショップ CANSec 2017 において、日本 PI の原が基調講演を行い、また、両国の博士課程の学生が本研究課題の成果の一部をポスター発表した。本研究課題を推進する上で貴重なイベントを主催できたことは、本共同研究の相乗効果と考える。

## 5-3 共同研究成果から期待される波及効果

本共同研究には、日本側から若手の助教（2 名）と両国側で博士課程の学生（5～6 名）が研究協力者として研究開発に関わった。この若手研究者らは、本共同研究を通じて急速に実力を伸ばし、世界最高峰の論文誌や国際会議に自身のキャリアで初めて採択されるなど、国際的にトップレベルの研究者へと成長した。本共同研究の経験を起点として、彼らの今後の益々の成長・活躍が期待される。

さらに、本研究では 5-2 で述べた主催イベントを通じて、異分野の研究者・技術者と人的ネットワークを構築した。この研究者ネットワークは、今後、災害とビッグデータに関する研究を、異分野の研究者の協働でさらに発展させる上で貴重な財産になる。

最後に、本研究で開発した二つのプラットフォームは、現状では一般公開するレベルまで成熟していないが、さらに研究開発を進め、両プラットフォーム間の連携機能を拡充することにより、センサ、IoT デバイス、SNS などのデータを透過的に扱うことができる重要な基盤になることが期待される。

Strategic International Collaborative Research Program (SICORP)  
Japan—U.S.A Joint Research Program  
Executive Summary of Final Report

1. Project Title : 「BDD: Efficient and Scalable Collection, Analytics and Processing of Big Data for Disaster Applications」
2. Project Period : April, 2015 ~ March, 2018
3. Main Participants :

Japan-side

	Name	Title	Affiliation	Role
PI	Takahiro Hara	Professor	Osaka University	Problem definition, Methodology design, System design/implementation
Collaborator	Shojiro Nishio	President	Osaka University	System implementation, Evaluation
Total number of participating researchers in the project: 2				

U.S.A -side

	Name	Title	Affiliation	Role
PI	Sanjay Madria	Professor	Missouri University of Science and Technology	Everything
Total number of participating researchers in the project: 1				

4. Summary of the joint project

In this joint research program, we have tried to achieve fundamental technologies for collecting, analyzing, and processing big data in disaster situations, which includes 4 topics; (1) compression of highly multi-dimensional data from multi-streams, (2) indexing and query processing for highly multi-dimensional data, (3) monitoring highly multi-dimensional data from multiple data sources, and (4) social sensing using microblog data.

We have tackled topic (1) as two work-packages (W1 and W4), where we have first developed a collaborative data compression mechanism by multiple nodes (e.g., sensors or clouds) for one-dimensional data stream (W1), and extended this mechanism for multi-dimensional data (W4). Each of the other three topics has been addressed as single work-package (W2, W3, W5). We have also worked on development of platforms to integrate our proposed mechanisms. W1, W2, and W4 have been taken the lead by the US team, and W3 and W5 by the Japan team. For W6, both teams have closely collaborated.

- W1: Compression of Multiple Sensor Streams of One-dimensional Data
- W2: Indexing of Highly Multi-dimensional Data
- W3: Monitoring of Highly Multi-dimensional Data Generated from Multiple Data sources
- W4: Compression of Multiple Sensor Streams of Highly Multi-dimensional Data
- W5: Microblog Data Analysis for Social Sensing Applications
- W6: Development of Platforms to Integrate All mechanisms

5. Outcomes of the joint project

5-1 Intellectual Merit

In WP1 and WP4, we have developed mechanisms in which multiple nodes (data sources) collaboratively compress stream data. In WP2, we have developed an indexing mechanism in which multiple nodes cooperate to construct an index and process some

typical queries on it. In WP3, we have developed various query processing techniques for monitoring big data streams from multiple data sources. These techniques achieved several tens to hundreds times faster processing time than existing approaches. In WP5, to utilize microblog data as social sensors, we have developed some social sensing techniques to extract various useful knowledge from microblog data, and some fundamental technologies which can be commonly used for various different social sensing applications. In WP6, we have developed two platforms to integrate all techniques and mechanisms proposed in our project.

In WP1 to WP5, we could successfully complete all tasks as planned in the proposal by properly keeping the independence of each work-package in each team, and frequently exchanging opinions. In WP6, while we have not completed the implementation yet, we plan to continue to work on this task after this program.

In total, we have published 17 journal papers, 26 international conference papers, and 32 domestic conference papers, and given 11 invited talks. These publications include many top-ranked journals (ACM TOIS, IEEE TKDE (2 papers), ACM TSAS, WWW) and conferences (ICDCS, EMNLP, EDBT, Big Data, SSDM). We have also received 6 research awards including Osaka Science Award.

### 5-2 Synergy through the Collaboration

Both PIs have high-level and complementary research expertise to conduct this joint program, and have collaboratively worked together in longer than 10 years before this program. Due to these merits, we could smoothly conduct this research program from the beginning to the end. As the result, we have developed high-level techniques and mechanisms which are required to cover a wide range of techniques in different fields. Some of our co-authored papers have been published in high-level conferences, and many papers by each team have been published in top-ranked journals and conferences. These achievements clearly show the synergic effect of this joint program.

Since both PIs have many experiences of system development, we could successfully develop platforms which can be used for big data analytics on sensor/IoT and SNS data.

We have also actively worked on event organization. In the beginning of this project, we organized an NII Shonan meeting on "Big Data and Disasters", in which 21 participants from different backgrounds (computer science, engineering, geography, and people movement analysis) attended from 4 countries. In the closing phase of this project, the US PI (Dr. Sanjay Madria) organized the international workshop (CANSec 2017), in which the Japan PI (Dr. Takahiro Hara) gave a keynote speech, and 2 PhD students from both teams presented their posters as outreach activities of this program. These activities which significantly contributed to our joint work also show the synergic effect of this joint program.

### 5-3 Potential Impacts on Society

Two young assistant professors and 5 to 6 PhD students have participated in this joint program to support our research. They have built good research experiences through this program, and have rapidly grown up as top-level international researchers, e.g., their papers have been accepted for the top-ranked journals and conferences. They are expected to further grow up to be leading researchers in the future.

Moreover, through this program, we organized two events where we have built excellent human networks. These networks will be valuable assets when advancing this program to more interdisciplinary project in the future.

Last but not least, while the two platforms which we have developed in this program are still not perfect and we cannot make them open to public now, we plan to further develop them by enhancing mechanisms to bridge the two platform. By doing so, we will be able to provide a powerful fundamental system which can transparently handle both sensor/IoT and SNS data.

## 共同研究における主要な研究成果リスト

### 1. 論文発表等

\* 原著論文（相手側研究チームとの共著論文）

\* 査読有り

1. X. Cao, S. Madria, and T. Hara. Efficient Z-order Encoding Based Multi-modal Data Compression in WSNs, Proc. Int. Conf. on Distributed Computing Systems, 2017 June; 2185-2191.

2. X. Cao, S. Madria, and T. Hara. A WSN Testbed for Z-order Encoding Based Multi-modal Data Compression. Proc. Int. Conf. on Sensing, Communication and Networking. 2017 June.

3. M. Yokoyama, T. Hara, and S. Madria. Efficient Diversified Set Monitoring for Mobile Sensor Stream Environments, Proc. IEEE Int. Conf. on Big Data. 2017 December; 500-507.

\* 原著論文（相手側研究チームを含まない日本側研究チームの論文）

発表件数：国際 28 件、日本国内 18 件、合計 46 件

\* 査読有り

1. D. Amagata and T. Hara. Monitoring MaxRS in Spatial Data Streams, Proc. Int. Conf. on Extending Database Technology. 2016 March; 317-328.

2. D. Amagata, Y. Sasaki, T. Hara, and S. Nishio. Probabilistic Nearest Neighbor Query Processing on Distributed Uncertain Data. Distributed and Parallel Databases. 2016 June; 34(2), 259-287.

3. M. Yokoyama and T. Hara. Efficient Top-k Result Diversification for Mobile Sensor Data, IEEE Int. Conf. on Distributed Computing Systems, 2016 June; 477-486.

4. D. Amagata, T. Hara, and S. Nishio. Sliding Window Top-k Dominating Query Processing over Distributed Data Streams. Distributed and Parallel Databases. 2016 December; 34(4), 535-566.

5. T. Yoshihisa and T. Hara. A Low-load Streams Processing Scheme for IoT Environments. Proc. Int. Conf. on Big Data, 2016 December; 263-272.

6. K. Udomlamlert, T. Hara, and S. Nishio. Subscription-based Data Aggregation Techniques for Top-k Monitoring Queries. World Wide Web. 2017 March; 20(2), 237-265.

7. D. Amagata and T. Hara. A General Framework for MaxRS and MaxCRS Monitoring in Spatial Data Streams, ACM Transactions on Spatial Algorithms and Systems. 2017 May; 3(1), 1-13.

8. M. Shirakawa, T. Hara, and S. Nishio. IDF for N-Grams, ACM Transactions on Information Systems. 2017 June; 36(1) Article 5.

9. M. Shirakawa, T. Hara, and T. Maekawa. Never Abandon Minorities: Exhaustive Extraction of Bursty Phrases on Microblogs Using Set Cover Problem, Proc. Int. Conf. on Empirical Methods in Natural Language Processing. 2017 September; 2348-2357.

10. D. Amagata and T. Hara. Mining Top-k Co-Occurrence Patterns across Multiple Streams, IEEE Transactions on Data Engineering. 2017 October; 29(10), 2249-2262.

11. D. Amagata, T. Hara, and M. Onizuka. Space Filling Approach for Distributed Processing of Top-k Dominating Queries, IEEE Transactions on Data Engineering. 2018 (to appear).

\*その他の著作物（相手側研究チームとの共著のみ）（総説、書籍など）

1. S. Madria and T. Hara. Big Data Analytics and Knowledge Discovery - 17th International Conference, DaWaK 2015, Proceedings. Lecture Notes in Computer Science 9263, Springer 2015. September 1-4 2015.

2. S. Madria and T. Hara. Big Data Analytics and Knowledge Discovery - 18th International Conference, DaWaK 2016, Proceedings. Lecture Notes in Computer Science 9829, Springer 2016. September 6-8 2016.

3. T. Hara, S. Sigg, L. Shu, F. De Pellegrini, C. Petrioli, S. Madria. IEEE Access Special Section Editorial: Emergent Topics for Mobile and Ubiquitous Systems in Smartphone, IOT, and Cloud Computing ERA, IEEE Access. 2017; 5: 27827-27830.

\*その他の著作物（相手側研究チームを含まない日本側研究チームの総説、書籍など）

1. 原 隆浩. センサ・モバイルビッグデータを有効利用するためのプラットフォーム技術. 情報処理. 2015 年 10 月; 56(10): 973-977.

2. 神崎 映光, 原 隆浩, 義久 智樹, 西尾 章治郎. モバイルセンサネットワークにおけるデータ処理技術～動きまわるセンサが拓く未来～. 情報処理. 2015 年 11 月; 56(11):1096-1101.

3. T. Hara. Fusion of Heterogeneous Mobile Data, Challenges and Solutions, Adaptive Mobile Computing: Advances in Processing Mobile Data Sets (Chapter 3), 2017 August; 47-63.

## 2. 学会発表

\*口頭発表（相手側研究チームとの連名発表）

発表件数：計 1 件（うち招待講演：計 1 件）

\*口頭発表（相手側研究チームを含まない日本側研究チームの発表）

発表件数：計 25 件（うち招待講演：計 5 件）

\*ポスター発表（相手側研究チームとの連名発表）

発表件数：計 1 件

\*ポスター発表（相手側研究チームを含まない日本側研究チームの発表）

発表件数：計 1 件

## 3. 主催したワークショップ・セミナー・シンポジウム等の開催

1. NII 湘南会議：Big Data: Challenges and Opportunities for Disaster Recovery、主催者：

Sanjay Madria（ミズーリ科学技術大学・教授）、原 隆浩（大阪大学・教授）、Cyrus Shahabi（南カリフォルニア大学・教授）、Calton Pu（ジョージア工科大学）、湘南国際村センター、神奈川、日本、2016 年 3 月 27 日～31 日、参加人数 21 名

2. Central Area Networking and Security Workshop (CANSec 2017)、主催者：Sanjay Madria（ミズーリ科学技術大学・教授）、Egemen K. Cetinkaya（ミズーリ科学技術大学・助教）、Bharat Bhargava（パデュー大学・教授）、Missouri University of Science and Technology、Rolla、MO、USA、2017 年 10 月 28 日～29 日、参加人数 100 名程

#### 4. 研究交流の実績

##### 【合同ミーティング】

- ・2016 年 3 月 26 日～2016 年 4 月 1 日：Sanjay Madria 教授との研究打合せ、湘南国際村センター、神奈川、日本
- ・2016 年 5 月 10 日～2016 年 5 月 14 日：Sanjay Madria 教授との研究打合せ、東京大学、東京および大阪大学、大阪、日本
- ・2016 年 6 月 27 日～2016 年 7 月 1 日：Sanjay Madria 教授との研究打合せ、奈良ホテル、奈良、日本
- ・2017 年 10 月 25 日～2017 年 10 月 30 日：Sanjay Madria 教授との研究打合せ（日本から学生 1 名を派遣）、ミズーリ科学技術大学、ローラ、アメリカ
- ・2017 年 12 月 19 日～2017 年 12 月 21 日：Sanjay Madria 教授との研究打合せ、コクヨホール、東京、日本
- ・両国のチームメンバーを交えて skype ミーティングを月 1 回程度の頻度で開催した。

#### 5. 特許出願

研究期間累積出願件数：0 件

#### 6. 受賞・新聞報道等

##### \*受賞

1. 優秀論文賞（報処理学会マルチメディア，分散，協調とモバイルシンポジウム 2015）、中山 侑紀、天方 大地、原 隆浩、西尾 章治郎、2015 年 8 月 24 日
2. 優秀論文賞（報処理学会マルチメディア，分散，協調とモバイルシンポジウム 2016）、松尾 和哉、後藤 啓介、神崎 映光、原 隆浩、2016 年 8 月 25 日
3. 優秀論文賞（報処理学会マルチメディア，分散，協調とモバイルシンポジウム 2016）、林田 秀平、天方 大地、原 隆浩、Xing Xie、2016 年 8 月 25 日
4. Best Paper Award (Advanced Mobile Computing & Multimedia), Kazuya Matsuo, Akimitsu Kanzaki, Takahiro Hara, 2016 年 11 月 29 日
5. 優秀論文賞（FIT2017 情報科学技術フォーラム）、天方 大地、原 隆浩、2017 年 12 月 1 日
6. 大阪科学賞、原 隆浩、2017 年 11 月 15 日

##### \*報道

1. NHK テレビ、「ニュース 8 45～関西のニュースと気象情報（20:45～21:00）」、「大阪科

学賞 大阪大学の原栄誉教授ら 2 人選出」、2017 年 9 月 14 日

2. 朝日新聞、社会 14 版（37 ページ）「大阪科学賞に熊谷・原教授」、2017 年 9 月 15 日
3. 読売新聞、社会 13S（31 ページ）「大阪科学賞に京大教授ら 2 人」、2017 年 9 月 15 日
4. 毎日新聞、社会 13 版（28 ページ）「熊谷氏、原氏に大阪科学賞」、2017 年 9 月 15 日

## 7. その他

特になし