

戦略的創造研究推進事業
－CREST・さきがけ複合領域－

研究領域

「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」

複合領域中間評価用資料

研究総括　： 喜連川 優
副研究総括： 柴山 悦哉

2018 年 3 月

目次

1. 研究領域の概要	1
(1) 戦略目標	1
(2) 研究領域	5
(3) 研究総括および副研究総括	6
(4) 採択課題・研究費	7
2. 研究総括のねらい	11
(1) 重点的に取り組むべきコア分野の設定	11
(2) 研究用データの整備	12
(3) 分野を超えた連携	12
(4) 国際連携	13
3. 研究課題選考について	13
(1) 選考方針	13
(2) さきがけの選考過程と選考結果	13
(3) CRESTの選考過程と選考結果	14
4. 領域アドバイザーについて	16
(1) さきがけ	16
(2) CREST	17
5. 研究領域の運営について	18
(1) 研究総括の運営方針	18
(2) 領域内外の連携の推進	19
(3) 研究進捗状況の把握と評価、それに基づく指導	21
(4) 研究費配分上の工夫	22
(5) 人材育成	23
6. 研究の経過と所見	24
(1) 研究総括のねらいに対する研究の状況	24
(2) 研究領域全体として見た時の特筆すべき成果の見通し	24
(3) 科学技術の進歩に資する研究成果	25
(4) 科学技術イノベーションに貢献する成果	27
7. 総合所見	28

1. 研究領域の概要

(1) 戦略目標

「分野を超えたビッグデータ利活用により新たな知識や洞察を得るための革新的な情報技術及びそれらを支える数理的手法の創出・高度化・体系化」

①戦略目標の内容

(i) 達成目標

情報科学・数理科学分野とビッグデータの利活用により大きな社会的インパクトを生むような様々な研究分野（アプリケーション分野）との協働により研究を進め、アプリケーション分野での課題解決を通じてビッグデータから新たな知識や洞察を得ることを可能とする次世代アプリケーション技術を創出し、高度化すると同時に、様々な分野のビッグデータを統合解析することを可能とする共通基盤技術の構築を目指す。そのため、以下の目標の達成を目指す。

○各アプリケーション分野においてビッグデータの利活用を推進しつつ様々な分野に展開することを想定した次世代アプリケーション基盤技術の創出・高度化

○様々な分野のビッグデータの統合解析を行うための次世代基盤技術の創出・高度化・体系化

(ii) 将来実現し得る重要課題の達成ビジョン

本戦略目標を実施し、「(i) 達成目標」に記載した研究成果が得られることで、様々な分野のビッグデータを統合解析するための共通基盤技術を構築することができ、分野を超えたビッグデータの利活用を実現することができる。構築された技術を用いることで、ビッグデータの利活用が有効な研究分野の論文データ、実験・シミュレーションデータ、観測データ等の高度利用が可能となり、社会科学・人文科学等を含む複数の分野が連携した異分野融合領域のイノベーション創出を加速させることができる。

本事業終了後、アカデミア・企業等が様々な分野のビッグデータを統合解析できる共通基盤技術を活用して、研究開発や実用化を推進する。これによって、イノベーションによる新産業・新市場の創出や、国際競争力の強化を推進し、「我が国の産業競争力の強化」、「研究情報基盤の整備」の達成に貢献することを目指す。

(iii) 具体的内容

(背景)

高度情報化社会の進展に伴い、デジタルデータが爆発的に増大するビッグデータ（情報爆発）時代が到来した。世界のデジタルデータの量は、民間調査機関の推計によれば、2020年には、約40ゼタバイト（2010年度時の約50倍）へ拡大する見込みである。

その質的・量的に膨大なデータ（ビッグデータ）には新たな知識や洞察を得られる可能性があるが、様々なデータを組み合わせて、大規模な処理を実行しようとする、想定外のデータや正常に分析できないデータが大きくなることが多く、現況においてはその多くのデータが整理・構造化されておらず、有効に活用できていない状況である。

このため、ビッグデータを効果的・効率的に収集・集約し、革新的な科学的手法により知識発見や新たな価値を創造することの重要性が、国際的に認識されてきている。第一の科学的手法である経験科学（実験）、第二の科学的手法である理論科学、第三の科学的手法である計算科学（シミュレーション）と並び、データ科学（data centric science = e-サイエンス）は第四の科学的手法と言われ、ビッグデータ時代における科学の新たな地平を拓（ひら）く方法論として注目されている。

（研究内容）

本戦略目標では、ビッグデータの解析を円滑に実行するための革新的な方法論等の創出等のため、2つの達成目標の実現を目指す。具体的には以下の研究を想定する。

(a) 各アプリケーション分野においてビッグデータの利活用を推進しつつ様々な分野に展開することを想定した次世代アプリケーション基盤技術の創出・高度化

個別のアプリケーション分野の課題解決とともに、固有技術の他分野展開や新規基盤要素技術の導入を強力に推進する。

- ・多様かつ大量のアプリケーションデータ（健康・医療データ、地球観測データ、防災関連データ、ソーシャルデータ等）の転送、圧縮、保管等を容易に実現するための研究

- ・画像データや3次元データ等の多様なデータを検索、比較、解析等することで有意な情報を抽出するための研究

- ・アプリケーションデータから新たな課題の発見や洞察をより正確に行うための研究（疾患要因の解明、気候変動予測、リアルタイム解析による減災、人のニーズの予測等）

- ・定量データから生体、自然現象等に係る多様な数理モデルを構築し、実測データと組み合わせることで新たな知見を得るような、発見的探索スタイルの研究アプローチ推進のための研究基盤創出

(b) 様々な分野のビッグデータの統合解析を行うための次世代基盤技術の創出・高度化・体系化

情報科学・数理科学分野や人文科学の研究者による、独自の新規基盤要素技術の創出や複数のアプリケーション分野に展開する新規要素技術の創出を行う。

- ・データクレンジング技術（ノイズ除去、データの正規化、不要なデータ変動の吸収等）やデータに対して自動的に意味や内容に係る注釈を付与する技術

- ・高度な圧縮技術、圧縮したままで検索する技術、秘密性や匿名性を損なわないままマイニ

ングする技術

- ・データマイニング技術や機械学習の高度化（大量・多様なデータからのモデリング技術、異種データから関連性を探索する技術等）
- ・多様なアプリケーションデータの相関や関係性から新たな洞察を導くための可視化技術
- ・ビッグデータを共有・流通するためのシステム技術（データの加工、メタデータ管理、トレーサビリティ、匿名化、セキュリティ、課金等）
- ・課題の本質やビッグデータの構造を見いだすための数理的手法

なお、(a) の次世代アプリケーション基盤技術の創出・高度化に当たっては、(b) の研究で得られる次世代基盤技術を取り込みながら推進することが効果的であり、また、(b) の次世代基盤技術の創出・高度化・体系化に当たっては、(a) の研究で得られる次世代アプリケーション基盤技術やデータを共有、活用しながら研究を進めることが効果的であることから、(a) と (b) の研究が相互に連携することが求められる。

(iv) 政策上の位置付け（政策体系における位置付け、政策上の必要性・緊急性等）

第4期科学技術基本計画では、「我が国が直面する重要課題への対応」において、「我が国の産業競争力の強化」として、電子デバイスや情報通信の利用、活用を支える基盤技術等、革新的な共通基盤技術に関する研究開発を推進するとともに、これらの技術の適切なオープン化戦略を促進すると掲げている。また、「科学技術の共通基盤の充実、強化」として、シミュレーションや e-サイエンス等の高度情報通信技術、数理科学等、複数領域に横断的に活用することが可能な科学技術や融合領域の科学技術に関する研究開発を推進すると掲げている。さらに、「国際水準の研究環境及び基盤の形成」において、「研究情報基盤の整備」として、研究情報基盤の強化に向けた取組を推進するため、研究情報全体を統合して検索、抽出することが可能な「知識インフラ」としてのシステムを構築し、展開すると掲げている。

文部科学省では、全国の大学等の研究者が、サイエンスに活用できる多分野にわたるデータ、情報、研究資料等を、オンラインにより、手軽に利用でき、最新の「データ科学」の手法を用いて、科学的あるいは社会的意義のある研究成果を得ることのできる「アカデミッククラウド環境」について、必要な議論、検討等を進めるため、研究振興局長の下に「アカデミッククラウドに関する検討会」を設置し、2012年4月から6月に、「データベース等の連携」、「システム環境の構築」、「データ科学の高度化に資する研究開発」の3点を検討課題として議論を行い、7月に提言「ビッグデータ時代におけるアカデミアの挑戦」において、ビッグデータに関する共通基盤技術の研究開発として、ビッグデータ処理の各段階（データ収集、蓄積・構造化、分析・処理、可視化）における基盤技術の研究開発等が必要との方向性を取りまとめた。

(v) 科学的裏付け（海外の研究動向を踏まえた必要性・緊急性・実現可能性等）

米国においては、2011年に科学技術に関する大統領諮問委員会（PCAST）が、連邦政府はビッグデータ技術への投資が少ないと結論づけたことに対応し、科学技術政策局（OSTP）が2012年3月29日にビッグデータイニシアチブに関する公告を発表した。このイニシアチブには6機関（NSF、NIH、DOD、DARPA、DOE、USGS）が総額2億ドルを投資し、データへのアクセス、体系化、知見を集める技術を改善、強化としている。欧州、アジアにおいても、ビッグデータに対する研究投資を実施しており、今後、激しい国際競争が予想される。具体的には、欧州では2020年までにICTにおける研究開発への公共支出を55億ユーロから110億ユーロへと倍増させ、大規模なパイロットプロジェクトを実施し、公共に利益のある分野における革新的かつ相互運用可能なソリューション（エネルギーや資源を節約するためのICT、持続可能な保険医療、電子政府、インテリジェント輸送システム等）を開発することとしている。

(vi) 検討の経緯

文部科学省の研究振興局長の下に設置したアカデミッククラウドに関する検討会においては、2012年7月4日に提言「ビッグデータ時代におけるアカデミアの挑戦」を取りまとめ、ビッグデータに関する共通基盤技術の研究開発として、ビッグデータ処理の各段階（データ収集、蓄積・構造化、分析・処理、可視化）における基盤技術の研究開発等が必要との方向性や具体的な研究開発事項について取りまとめた。

これを踏まえ、科学技術・学術審議会研究計画・評価分科会情報科学技術委員会（第77回、第78回）（2012年7月5日、8月2日）においても、様々な分野における知的活動の成果として生み出されている大量データを効果的・効率的に収集・集約し、革新的な科学的手法により情報処理を行うことにより、新たな知的価値を創造する「データ科学」が重要との共通認識のもと、ビッグデータを利活用するための共通基盤技術の研究開発が必要との見解が示された。

本戦略目標は、これらの検討の結果を踏まえて作成したものである。

②戦略目標設定時以降の政策変化

(i) 第5期科学技術計画（2016年度～2020年度）

内閣総理大臣からの諮問を受け、総合科学技術・イノベーション会議は、第5期科学技術基本計画（2016～2020年度）の策定に向けた検討の結果を「科学技術基本計画について」として取りまとめ、2015年12月に答申を行った。その後、2016年1月に第5期科学技術基本計画が閣議決定され、本研究領域の発足から3年後の2016年度よりこの第5期基本計画に沿った新たな科学技術政策が推進されている。

第5期科学技術基本計画においては、サイバー空間とフィジカル空間（現実社会）が高度

に融合した「超スマート社会」を未来の姿として共有し、その実現に向けた一連の取り組みを「Society 5.0」と呼んでいる。そして、超スマート社会サービスプラットフォームの構築に必要な基盤技術の一つに、本研究領域の主たるテーマであるビッグデータ解析技術をあげ、本研究領域とも密接に関係するサイバーセキュリティ技術、IoTシステム構築技術、AI技術などとともに、国が速やかな強化を図ることとしている。

(ii) AIPプロジェクト (2016年度～)

文部科学省は、2016年度にAIPプロジェクト(人工知能/ビッグデータ/IoT/サイバーセキュリティ統合プロジェクト)を開始した。これは世界最先端の人材を結集し、革新的な人工知能技術の中核として、ビッグデータ・IoT・サイバーセキュリティを統合した研究開発を行う拠点の新設や、イノベーションを切り開く独創的な研究者等の支援を推進することを目的にしたものである。AIPプロジェクトの実施機関として、理化学研究所に革新知能統合研究センター(通称AIPセンター)が、JSTに戦略的創造研究推進事業の関連する研究領域で編成した「AIPネットワークラボ」が設置され、AIPプロジェクトの推進にあたっては、両機関が連携して取り組んでいる。「AIPネットワークラボ」には関連するCREST、さきがけ、ACT-Iの領域が含まれ、本研究領域も属している。「AIPネットワークラボ」の以下の4つの活動方針に沿った施策に本研究領域も参加し、成果の拡大に貢献している。

- AI関連分野の研究をさらに先導し、存在感を発揮する
- 国内外に積極的に研究成果を発信し、研究分野の進展に貢献する
- ラボ内の共同研究等を支援し、新たな価値を創造する
- 若手研究者の育成と教育に、ラボ全体で取り組む

(2) 研究領域

「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」

(2013年度発足)

本領域は、戦略目標の2つの達成目標(1①(i)参照)のうち「様々な分野のビッグデータの統合解析を行うための次世代基盤技術の創出・高度化・体系化」の達成を目指すものである。

具体的には、機械学習やデータマイニングの高度化(大量・多様なデータからのモデリング、統計処理の高度化、ストリーミング処理、圧縮、ノイズ除去などを含む)、大量データの処理を可能とする高性能アルゴリズムと高性能システムアーキテクチャ、自然言語データや画像データからの有意な情報の抽出、ビッグデータを守るセキュリティ技術とプライバシー保護技術の高度化などの研究開発を推進している。これらの研究の推進にあたり、ビッグデータから社会における価値創造に至るシステム全体の設計を視野に入れ、

ICT 以外の分野との積極的な連携によって社会受容性の高い次世代共通基盤技術の創出・高度化・体系化に取り組む。

また、関連領域の「科学的発見・社会的課題解決に向けた各分野のビッグデータ利活用推進のための次世代アプリケーション技術の創出・高度化」に基盤技術を提供するとともに、同領域で得られる次世代アプリケーション技術やデータを共有・活用するなどの連携を推進する。

(3) 研究総括および副研究総括

(所属・役職は中間評価時点)

研究総括：

氏名 喜連川 優 (所属 国立情報学研究所 役職 所長／
所属 東京大学生産技術研究所 役職 教授)

副研究総括：

氏名 柴山 悦哉 (所属 東京大学情報基盤センター 役職 教授)

(4) 採択課題・研究費

① さきがけ

(百万円)

採択年度	研究代表者	中間評価時 所属・役職	研究課題	研究費*
2013 年度	生貝直人	株式会社 情報通 信総合研究所	ビッグデータ統合利活用の ための法制度のあり方に関 わる総合的日米欧比較研究 による制度設計	8
	佐藤一誠	東京大学 講師	統計的潜在意味解析による データ駆動インテリジェン スの創発	40
	田部井靖生	理化学研究所革新 知能統合研究セン ター ユニットリ ーダ	透過的データ圧縮による高 速かつ省メモリなビッグデ ータ活用技術の創出	29
	松谷宏紀	慶應義塾大学 准 教授	多様な構造型ストレージ技 術を統合可能な再構成可能 データベース技術	38
	水野貴之	国立情報学研究所 准教授	金融ビッグデータによるバ ブルの早期警戒技術の創出	40
	宮尾祐介	国立情報学研究所 准教授	非テキストデータと接続可 能なテキスト解析・推論技 術の開発	38
2014 年度	大竹義人	奈良先端科学技術 大学院大学 准教 授	統計学習と生体シミュレ ーションを融合した循環型手 術支援	40
	河原大輔	京都大学 准教授	計算機・人の知を統合した ビッグテキスト解析基盤	37
	杉山磨人	国立情報学研究所 准教授	統計的有意性を担保する超 高速パターン発見技術の創 出	41
	山本泰生	山梨大学 助教	高次知識を獲得するリソー ス指向型オンラインマイニ	28

			ング法の開発	
2015 年度	秋葉拓哉	国立情報学研究所 特任助教	大規模複雑データのインタラクティブ解析を可能にする高レスポンス情報抽出技術の開発	6
	小野木章雄	東京大学 農学研究員 (さきがけ専任研究員)	膨大なレガシー栽培データを蘇生する (データさきがけ)	30
	酒向重行	東京大学 助教	タイムドメイン宇宙観測用動画データの高速逐次処理法の開発	38
	佐藤彰洋	京都大学 特定准教授	グローバル・システムの持続可能性評価基盤に関する研究	35
	島田敬士	九州大学 准教授	時空間粒度の異なる教育ビッグデータの非同期ストリーム処理基盤の構築	40
	柳澤琢史	大阪大学 寄附研究部門講師	思考・行動を予想する脳ビッグデータ (データさきがけ)	32
	山田拓司	東京工業大学 准教授	ヒト腸内環境ビッグデータ (データさきがけ)	36
			総研究費	556

*研究費：2017 年度上期までの実績額に 2017 年度下期以降の計画額を加算した金額

さきがけ：重点配分した趣旨

当初配分では、計画実施のための必要性のみに基づいた配分を行っており、重点的な配分は行っていない。課題採択後には、さきがけ研究者が、領域内外の CREST・さきがけの研究者と連携して共同研究等を行う活動に対し、増額を行った。ビッグデータ基盤技術の研究では、アプリケーション分野の実データを用いることで新たな展望が開けることが珍しくないが、一方で、基盤技術の研究とアプリケーション分野のデータ取得を 1 人のさきがけ研究者が同時に行うのは容易でない。そこで、アプリケーション分野のデータを利用できる研究者との連携、後述の「データさきがけ」の研究者が整備したデータを本研究領域内の基盤技術の研究者に提供するための連携を推進するため、3 名に計 9 百万円の重点的な増額を行った。

②CREST

(百万円)

採択年度	研究代表者	中間評価時 所属・役職	研究課題	研究費*
2013年度	黒橋禎夫	京都大学 教授	知識に基づく構造的言語処理の確立と知識インフラの構築	407
	佐久間淳	筑波大学 教授	自己情報コントロール機構を持つプライバシー保護データ収集・解析基盤の構築と個別化医療・ゲノム疫学への展開	273
	松岡聡	東京工業大学 教授	EBD: 次世代の年ヨッタバイト処理に向けたエクストリームビッグデータの基盤技術	381
	山西健司	東京大学 教授	複雑データからのディープナレッジ発見と価値化	333
2014年度	宇野毅明	国立情報学研究所 教授	データ粒子化による高速高精度な次世代マイニング技術の創出	301
	加藤直樹	関西学院大学 教授	ビッグデータ時代に向けた革新的アルゴリズム基盤	292
	原田達也	東京大学 教授	膨大なマルチメディアデータの理解・要約・検索基盤の構築	403
	宮地充子	大阪大学 教授	ビッグデータ統合利活用促進のためのセキュリティ基盤技術の体系化	391
2015年度	合田憲人	国立情報学研究所 教授	インタークラウドを活用したアプリケーション中心型オーバーレイクラウド技術に関する研究	315
	津田宏治	東京大学 教授	離散構造統計学の創出と癌科学への展開	300
	山名早人	早稲田大学 教授	ビッグデータ統合利用のためのセキュアなコンテンツ共有・流通基盤の構築	291
			総研究費	3,687

*研究費：2017年度上期までの実績額に2017年度下期以降の計画額を加算した金額

CREST：重点配分した趣旨

当初配分では、計画実施のための必要性を重視しつつ、「モノへの投資より人とデータへの投資」(5(4)参照)という考え方に基づいた配分を行った。総研究費 3,687 百万円の約 50% が人件費・謝金に充当されており、設備備品費は 10%未満である。課題採択後には、成果の実用化に向けた実証実験の準備・推進と成果のオープンソース化、NSF を含む領域内外の研究者との連携の推進、領域発足以降の深層学習の急速な進歩などに起因する計画の拡充に対し、重点的に追加配分を行った(2016 年度以降は、AIP ネットワークラボ特別費用からの追加配分)。

2. 研究総括のねらい

アマゾン、グーグル、フェイスブックに代表されるグローバルメガサービスが、大量データ処理基盤とサービスプラットフォームを構築し、主にビジネスの分野でビッグデータ時代を牽引してきた。一方、高エネルギー物理学、ゲノム科学、天文学を始めとしたサイエンスにおいても、大規模データが科学的発見に重要な役割を果たすことが認識され、第4期科学技術基本計画でも、大規模データに基づくサイエンスは第4の科学（e-サイエンス）と位置づけられた。続く第5期科学技術基本計画では、世界に先駆けた「超スマート社会」の実現（Society 5.0）が提唱され、ビッグデータ解析は、超スマート社会サービスプラットフォームに必要となる基盤技術と位置づけられている。

このようにビジネス、サイエンス、社会の活動が大規模なデータに駆動されるビッグデータ時代においては、データ基盤・サービスプラットフォームこそがこれらの活動を支える生命線となる。そこで、本領域では、将来の科学・産業・社会を支えるデータ基盤・サービスプラットフォームの構築に資する技術の研究開発を推進する。さらに、研究のために不可欠なデータの整備にも取り組む。これらを実現するため、以下の(1)～(4)のねらいを定めた。

(1) 重点的に取り組むべきコア分野の設定

ビッグデータの関連領域は広大であり、基盤技術に限定しても数多の課題が山積している。そこで、次の(i)～(iv)の4つのコア技術分野と(v)の技術とは補完的な法制度の分野を設定し、研究分野の重点化を図った。

- (i) 機械学習、統計解析、データマイニングなどのビッグデータ解析において必須となる数理的な手法やアルゴリズムを開発する数理基礎技術分野
- (ii) ビッグデータ時代に大量に取得可能となる自然言語や画像のデータを解析し、高度な情報・意味の抽出を目指す自然言語・画像基盤技術分野
- (iii) ビッグデータをサイバー攻撃や事故から守り、プライバシー情報の漏洩等を防ぐ技術を開発するセキュリティ基盤技術分野
- (iv) 大量かつ多様なビッグデータを効率的に処理するデータ基盤・サービスプラットフォームの構成技術を開発するシステムアーキテクチャ基盤技術分野
- (v) ビッグデータ時代における制度面での検討を行うビッグデータ法制度分野

これらの基礎・基盤技術と制度は、将来のビッグデータプラットフォームを構築し、社会の中で有効に活用するために必須のものである。

(i)は、特定のアプリケーション分野に依存せず、大量で多様なビッグデータから、意味のある情報を抽出するための基礎技術である。ビッグデータ解析のアプリケーション分野は非常に多岐に渡ることが予想されるため、多様なアプリケーション分野へ展開可能な汎用的で核となる技術を持つことは、将来のためにきわめて重要である。

(ii)は、長年に渡り人類が知識を蓄積するために用いて来た言語や画像というメディア

のビッグデータから、機械が高次の意味を読み取るための基盤技術である。単にデータの量が多いだけでなく、埋もれている有意義な情報や知識が多いことが期待できる。スモールデータに限定すれば人間には簡単に読み取れる意味が、従来、機械にはなかなか読み取れなかったメディアであるが、近年、深層学習の発達等により技術レベルが急激に向上している。

(iii)は、ビッグデータプラットフォームが有するデータの保護のために必要なものである。ビッグデータプラットフォームの社会的有用性が高くなればなるほど、個人、組織、国家等にとってデータ保護の重要性も高くなる。また、高いレベルの保護が実現できればできるほど、より機微なデータまで扱うことが妥当となり、社会的有用性も向上する。逆にビッグデータを保護する技術の研究が遅れると、どんなに優れた解析技術を開発しても、実用上は使えない状況に陥りかねない。後付で泥縄の対応が行われることがないように、他の技術開発と並行して推進する必要がある。

(iv)は、ビッグデータの取得・蓄積・解析等を行うシステムの構成法に関する基盤技術である。大量のプロセッサ、メモリ、ストレージ、ネットワーク等を構成要素とし、地理的に分散した高性能・高信頼システムを構築する技術を開発するとともに、そのシステムの機能を、(i)～(iii)のコア技術を用いてビッグデータ解析等を行う利用者、あるいはそのための開発・運用を行う利用者に、使いやすい形で提供するための研究も必要になる。

最後の(v)は、ビッグデータの利活用を推進すると同時に適切な保護も行い、社会が受容可能なバランスの取れた技術の利用方法を検討するためのものである。法制度整備の検討も技術開発と並行して推進する必要がある。

(2) 研究用データの整備

ビッグデータ基盤技術の研究を推進するにあたり、研究に利用可能なビッグデータの整備は、きわめて重要である。しかし、第三者が研究に利用可能なレベルのデータを大量に整備するためには大きな手間がかかること、手間をかけてデータを整備しても研究者としては評価されにくいことなどから、ビッグデータ基盤技術分野の研究者が自由に使えるデータの整備はあまり進んでいないのが実情である。このような現状に鑑み、研究用データの整備にも取り組む。その際に、分野を超えた統合解析技術の研究開発に役立てるためには、様々な種類やアプリケーション分野のデータを揃えることが望ましい点にも配慮する。

(3) 分野を超えた連携

本領域は、分野を超えたビッグデータ統合利活用のための共通基盤技術の構築を目指しており、様々なアプリケーション分野との連携が重要となる。また、戦略目標でも述べられている通り、「次世代基盤技術の創出・高度化・体系化」と「次世代アプリケーション基盤技術的の創出・高度化」をそれぞれに担当する本領域とCREST「科学的発見・社会的課題解決に向けた各分野のビッグデータ利活用推進のための次世代アプリケーション技術の創出・高度化」研究領域（以下、ビッグデータ応用領域）が相互連携することで、両領域の研

究を効果的に進めることが期待できる。そこで、本領域の研究開発を推進するにあたっては、ビッグデータ応用領域との連携・協働を重視する。

(4) 国際連携

ビッグデータに関する研究投資は国外でも活発に行われている。特に米国においては、2012年には総額2億ドルの投資を表明するビッグデータイニシアチブの公告も行われており、比較的早い時期からビッグデータに関連するプログラムが始まっている。これらのプログラムの研究者と連携することにより、研究の加速が期待できる。そこで、国際連携も重視し、米国NSFとシンポジウムの合同開催等を行い、人的交流を促進することで、共同研究を進めるための環境を整える。

3. 研究課題選考について

(1) 選考方針

本研究領域では、将来の科学・産業・社会を支えるデータ基盤・サービスプラットフォームの構築に資する技術の研究開発を推進し、さらに、研究のために不可欠なデータの整備にも取り組む。そのために、「数理基礎技術分野」「自然言語・画像基盤技術分野」「セキュリティ基盤技術分野」「システムアーキテクチャ基盤技術分野」の4つのコア技術分野と「ビッグデータ法制度分野」を設定した。そして、本研究領域全体で、これらの分野の主要テーマをカバーできるように課題を選考した。分野・専門に関しては、CREST、さきがけともにこの狙いに合致する研究課題を選考することにした。

さらに、研究構想に高い独創性・新規性が認められること、他とのコラボレーションが期待できること、また提案者が明確な目的意識を有していることを重視して審査を行った。また、さきがけ課題の審査では挑戦的であることをより重視し、CREST課題の審査ではイノベーションへの発展が期待できることをより重視した。特に、CREST課題については、基盤技術だけではなくその展開先まで含めた提案を高く評価した。

なお、さきがけにおいては2014年度から、アプリケーション分野のビッグデータの新規取得・準備やデータの他研究者への提供を目指すさきがけ研究提案（「データさきがけ」と呼ぶ）も募集対象とした。

(2) さきがけの選考過程と選考結果

2013年度から2015年度まで計3回の公募に対して、情報通信分野や数理科学分野はもちろん、ゲノム科学、脳科学、農学、医学などのライフサイエンス系、地球科学、天文学、物質材料科学などの自然科学系や工学系、法学、経済学、経営学、教育学、心理学、社会学などの人文社会系、さらに環境学や防災・減災などの総合系にわたる実にさまざま幅広い研究分野から、2013年度は100件、2014年度は35件、2015年度も35件の応募があった。こ

これらの研究提案を領域アドバイザー（2013年度は15名、2014年度と2015年度は13名）の協力を得て書類選考を行い、特に優れた研究提案として各年度14件を面接対象とした。書類選考では、きわめて高い評価を行う評価者が一人でもいることを重視し、評価が大きく割れる課題については、それぞれの立場の評価者の意見を詳しく吟味した。面接選考会でも評価が割れる課題を中心に議論を行った。採択課題数は2013年度が6件、2014年度が4件、2015年度が7件である。2015年度の採択課題には、「データさきがけ」の課題が3件含まれており、これらの課題には、領域全体の研究連携や加速展開、さらには領域外の研究者への貢献を期待した。

以上の17件のさきがけ採択課題は、

「数理基礎技術分野」：佐藤（一）、杉山、佐藤（彰）、秋葉、山本、島田、田部井

「自然言語・画像基盤技術分野」：宮尾、河原、大竹、酒向、水野

「システムアーキテクチャ基盤技術分野」松谷

「ビッグデータ法制度分野」生貝

の4つの技術・制度の分野に分布し、さらに、

「データさきがけ」：小野木、柳澤、山田

の取り組みも得られた。

これら採択者の中には、アプリケーション分野に軸足を置きながらも、分野を超えた基盤技術の研究や研究用データの整備を行う研究者が含まれている。具体的なアプリケーション分野としては、経済、教育、観光、医療、農業、天文など非常に多様である。コア分野として、自然言語や法制度の課題も採択しており、理工学やライフサイエンスはもちろん、人文社会系に関しても、一般的な大学の法学部、経済学部、文学部、教育学部に関係する分野にまたがっている。

以上のことから、広範な分野の研究者が、本研究領域に興味を持って応募し、採択された課題に限っても高い多様性が保たれていることがわかる。異なる分野の異なる発想をぶつけ合う場を醸成することが、イノベーション創出のために重要であり、この結果は特筆に値する。

(3) CRESTの選考過程と選考結果

2013年度から2015年度まで計3回の公募に対し、様々な用途の機械学習・解析技術、次世代システムアーキテクチャ、セキュリティ基盤技術、応用を想定したさまざまな基盤技術などの応募が、2013年度は50件、2014年度は20件、2015年度は21件あった。これらの研究提案を領域アドバイザー（2013年度は13名、2014年度と2015年度は11名）の協力を得て書類選考を行い、特に優れた研究提案として2013年度と2014年度は各12件、2015年度は10件を面接対象とした。さきがけの選考と同様に、書類選考では、きわめて高い評価を行う評価者が一人でもいることを重視し、評価が大きく割れる課題については、それぞれの立場の評価者の意見を詳しく吟味した。さきがけと異なる点としては、チーム内のシナジ

一を重視し、尖った基盤技術とその展開先が有効に結びついたチーム構成などを高く評価した。面接選考会でも評価が割れる課題を中心に議論を行った。そして、2013年度と2014年度は各4件、2015年度は3件の課題を採択した。これら計11件の採択課題と事前に設定した分野の関係は以下の通りである。事前に設定した5つの分野を網羅しており、いずれも高い実績を有し、社会的インパクトの大きな成果を期待できる研究者チームである。

(2013年度採択課題)

山西チーム：数理基礎技術分野

黒橋チーム：自然言語・画像基盤技術分野

佐久間チーム：セキュリティ基盤技術分野

松岡チーム：システムアーキテクチャ基盤技術

(2014年度採択課題)

宇野チーム：数理基礎技術分野

加藤チーム：数理基礎技術分野

原田チーム：自然言語・画像基礎技術分野、数理基礎技術分野

宮地チーム：セキュリティ基盤技術分野

(2015年度採択課題)

合田チーム：システムアーキテクチャ基盤技術分野

津田チーム：数理基礎技術分野

山名チーム：セキュリティ基盤技術分野、ビッグデータ法制度分野

なお、各課題の代表者は、大学または国立の研究機関に所属する情報分野の研究者であるが、主たる共同研究者として、2課題には企業の研究者、8課題にはアプリケーション分野の研究者が含まれている。

4. 領域アドバイザーについて

(1) さきがけ

(※退任者は、退任時の所属・役職を示す)

領域アドバイザー名	現在の所属 ※	役職 ※	任期
荒川 薫	明治大学	学部長 教授	2013年6月～2021年3月
石塚 満	東京大学	名誉教授	2013年6月～2021年3月
岩野 和生 ※	東京工業大学	客員教授	2013年6月～2015年10月
上田 修功	NTT	研究室長	2013年6月～2021年3月
柴山 悦哉 ※	東京大学	教授	2013年6月～2014年3月 (退任後、副研究総括に就任)
田中 英彦	東京大学	名誉教授	2013年6月～2021年3月
辻井 潤一	産業技術総合 研究所	センター長	2013年6月～2021年3月
徳田 英幸	情報通信研究 機構	理事長	2013年6月～2021年3月
徳山 豪	東北大学	教授	2013年6月～2021年3月
東野 輝夫	大阪大学	教授	2013年6月～2021年3月
堀 浩一	東京大学	教授	2017年12月～2021年3月
室田 一雄 ※	東京大学	教授	2013年6月～2015年9月
安浦 寛人 ※	九州大学	理事 副学長	2013年6月～2014年3月
北川 博之 (さきがけ専任)	筑波大学	教授	2013年6月～2021年3月
山西 健司 (さきがけ専任)	東京大学	教授	2013年6月～2021年3月
Calton Pu	Georgia Institute of Technology (米)	Professor	2013年6月～2021年3月
Nozha Boujemaa	Inria (仏)	Director of Research	2016年10月～2021年3月

(2) CREST

(※退任者は、退任時の所属・役職を示す。)

領域アドバイザー名	現在の所属 ※	役職 ※	任期
荒川 薫	明治大学	学部長 教授	2013年6月～2021年3月
石塚 満	東京大学	名誉教授	2013年6月～2021年3月
岩野 和生 ※	東京工業大学	客員教授	2013年6月～2015年10月
上田 修功	NTT	研究室長	2013年6月～2021年3月
柴山 悦哉 ※	東京大学	教授	2013年6月～2014年3月 (退任後、副研究総括に就任)
田中 英彦	東京大学	名誉教授	2013年6月～2021年3月
辻井 潤一	産業技術総合 研究所	センター長	2013年6月～2021年3月
徳田 英幸	情報通信研究 機構	理事長	2013年6月～2021年3月
徳山 豪	東北大学	教授	2013年6月～2021年3月
東野 輝夫	大阪大学	教授	2013年6月～2021年3月
堀 浩一	東京大学	教授	2017年12月～2021年3月
室田 一雄 ※	東京大学	教授	2013年6月～2015年9月
安浦 寛人 ※	九州大学	理事 副学長	2013年6月～2014年3月
Calton Pu	Georgia Institute of Technology (米)	Professor	2013年6月～2021年3月
Nozha Boujemaa	Inria (仏)	Director of Research	2016年10月～2021年3月

以下のような考え方で、領域アドバイザーの人選を行った。

- ① 複合領域の特徴を活かすために、さきがけ専任の2名以外は、全員 CREST/さきがけ兼任のアドバイザーとした。さきがけ専任アドバイザー2名には若手研究者への適切な助言

を期待できる研究者を選定した。

- ② 本研究領域の対象分野をおおよそカバーできる「システム」、「OS」、「データベース」、「セキュリティ」、「画像処理」、「AI」、「機械学習」、「自然言語処理」、「アルゴリズム」、「プログラミング言語」「センサ」「ネットワーク」などの多様な専門分野から、高い研究実績を有する研究者を選定した。
- ③ 戦略目標の達成ビジョン (1①(ii)参照) で述べられている「イノベーションによる新産業・新市場の創出」に向けた評価・助言を強化するため、国内アドバイザー15名（退任者を含む）のうち約半数の7名は、企業に勤務しているか勤務した経験のある研究者から選定した。この7名の中には、国際企業の海外の研究所での勤務経験がある研究者2名が含まれる。
- ④ 国際競争力強化と国際連携推進のために、米国の研究者を本研究領域発足直後に選定した。その後、欧州フランスの研究者も選定し、海外アドバイザーを充実させた。
- ⑤ 途中退任した国内アドバイザーを補充するために、2017年に国内アドバイザーを1名追加した。

領域アドバイザーからは、選考における評価・助言、それぞれの課題の研究に対する評価・助言、国際連携や将来のイノベーション創出に関する助言、さらには領域全体の運営に関わることまで、研究の細部に関するものから、大所高所のものまでの的確なアドバイスを頂いている。

5. 研究領域の運営について

(1) 研究総括の運営方針

各課題の研究の進展のためには、研究者の自由な発想が重要であり、それが最大限発揮できるようにすべきである。一方、「2. 研究総括のねらい」で述べた内容を着実に実施するためには、トップダウンの施策やアドバイスの実施が欠かせない。本領域の運営にあたっては、この両者のバランスを取ることを重視している。

各課題の研究活動の推進については、研究代表者および参画者が行う研究の実施状況を適宜確認し、必要に応じてアドバイスを行う体制としている。そのために、合同領域会議を開催し、原則として、CREST・さきがけの全課題の研究経過・成果に関する報告とそれに対するアドバイスを行っている。さらに、サイトビジット（CREST・さきがけ共通）、中間評価（CRESTのみ）、合宿型さきがけ領域会議（さきがけのみ）により、より詳細な研究経過・成果の確認とアドバイスを行っている。研究進捗状況の把握・評価と指導については、(3)で詳しく述べる。

「2. 研究総括のねらい」における連携を加速するために、領域内連携、ビッグデータ応用領域との連携、国際連携を加速する試みも行っている。領域内およびビッグデータ応用領域との連携のために、ビッグデータ応用領域との合同領域会議を年に1回開催している。ま

た、国際連携のために、海外のビッグデータ研究プログラムとのリエゾンになる海外アドバイザーの招聘、領域会議への海外研究者の招待、米国 NSF との公開シンポジウムの合同開催、日米の研究者のマッチングを目的としたワークショップの開催などを実施している。領域内外との連携の推進については、(2)で詳しく述べる。

基盤技術をイノベーション創出とその社会実装につなげるためには、個々の要素技術を磨くだけでは不十分である。実世界の価値観を理解し、データの取得や要素技術の統合を、社会制度を含む実世界の制約とすり合わせながら進めていく作業等が必要となる。特に CREST の研究者に対しては、このような過程を経験し、将来のインパクトが見込める研究を行うために、実証実験の促進を勧めている。

(2) 領域内外の連携の推進

①ビッグデータ応用領域やその他の領域との連携

ビッグデータ応用領域との連携は、年に1回のペースで開催する合同領域会議において、両領域の研究者間で情報の共有を促し、さらに共同研究を勧めるアドバイスをを行うことで推進している。このような試みから、基盤技術を有する本領域の研究者と応用技術・データを有するビッグデータ応用領域の研究者による共同研究が生まれている。代表的なものとしては、本領域さきがけ1期生の田部井が有するデータ圧縮技術をビッグデータ応用領域の CREST 船津チームが有する創薬のための化学物質のライブラリーに適用した研究をあげることができる。別のタイプの連携事例として、さきがけ1期生の佐藤や2期生の杉山は、生命科学分野のさきがけ領域の研究者との共同研究を開始するなど、積極的に他分野連携を進めている。

②領域内連携

本領域内部での連携についても、次のような点を考慮して、推進を図っている。

- (i) さきがけ研究は個人型であり、個人の力による尖った研究を進めるのが原則である。しかし、ビッグデータを活用して実世界の問題を解決するためには、広範な知見を要する可能性が高い。そこで、他の研究者の知見も積極的に活用可能とすべく、相互情報交換の場を設ける。
- (ii) CREST は、さきがけに比べると自己完結的だが、集中的に取り組むべき大きな問題を解決するために、相補的な関係にあるチーム間の連携を推進する。

これらのうち(i)については、領域間連携の場合と同様に領域内連携のための場を設け、連携を推進するようにアドバイスをを行なっている。ただし、具体的な共同研究の内容は、個々の研究者の自由な発想を尊重する方針である。意見交換の場としては、CREST と合同の領域会議の他に、さきがけ研究者のみを対象とした領域会議も開催している。これらの結果、さきがけ研究者同士や CREST 研究者とさきがけ研究者による共同研究が始まっている。前

者の事例としては、経済物理学、自然言語処理、統計学をそれぞれ専門とするさきがけ1期生の水野、宮尾と2期生の杉山が協力する形で共同研究を行っており、「さきがけネットワーク」にも応募して採択されている。また、後者の代表的事例としては、CREST 津田チームとさきがけ2期生の杉山が共同で、統計学の問題に情報幾何の概念を適用し、効率的な計算方法を導く研究を行った。

一方、(ii)のタイプの連携を進めるためには、問題を共有しうる課題間の連携を促進する工夫を行い、共通の問題に対しては協力して取り組むようにアドバイスを行なっている。この結果、セキュリティとプライバシー保護を主要なテーマとするCRESTの佐久間チーム、宮地チーム、山名チームの連携が進んでおり、2017年には3チーム共同による国際ワークショップを開催した。また、システムアーキテクチャに関する研究を進めるCRESTの松岡チームと合田チームの間では、広域分散ファイルシステムへの高速アクセスを実現するソフトウェアを松岡チームが開発し、合田チームが利用するなどの協力関係が形成されている。

③国際連携

米国NSFで行われている研究プロジェクトとの連携のための施策を進めており、今後はEUとの連携も推進したいと考えている。NSFとの連携を進めるための施策として、まず、米国ジョージア工科大学のCalton Pu教授をCREST・さきがけの領域アドバイザーとして招聘し、NSFとのリエゾンとしての役割を依頼した。Pu教授には、課題採択の審査時に、NSFでの経験を踏まえたアドバイスも頂いている。

そして、具体的な研究成果がある程度蓄積されるのを待ち、2016年度からNSFとの合同イベントの開催、非公開の会合への相互招待を開始した。合同領域会議へのNSF研究者の招待、NSFのビッグデータPI(Principal Investigator)ミーティングへのJST研究者の招待、公開シンポジウムの共同開催、国際会議でのワークショップの共同企画などを行っており、日米で各3回開催している(1回が2~3日間で、回によっては複数のイベントが併設されている)。NSF研究者を招待した領域会議では、発表や質疑をすべて英語で行い、有益な情報交換を目指している。NSFの研究施策や研究動向を知る研究者からのコメント等には、国際連携の推進にあたって有益なものがある。また、運営に関する意見交換も行っており、NSF側でも、Dear College Letter: National Science Foundation (NSF)-Japan Science and Technology Agency (JST) Collaborative Researchを2017年に発行し、本研究領域およびビッグデータ応用領域の研究者との連携をより強く支援する体制を築くに至った。

NSFとの合同イベントの中には、日米の研究グループのマッチングを行うことを目的としたものも含まれており、2017年に日米で各1回開催している。マッチングを効果的に行うため、日米で共同研究を行う可能性があるグループのペアを事前に抽出している。この作業を行うにあたっては、JSTとNSFの両側の状況を知悉したPu教授に大きく貢献して頂いている。現在までに、本研究領域とNSFの研究者の間で8件の共同研究が始まりつつある。例えば、CREST 山名チームと米国 ミズーリ工科大学 Sajal K. Das 教授グループ(NSF

projects (CNS-1545037 and CNS-1545050))との共同研究は、2016年度の交流で始まり、2017年度に大きく進捗した。「サイバーフィジカルアプリケーションにおけるビッグデータとIoTセキュリティ (BITS: Big Data and IoT Security in CPS Applications)」を共同研究テーマとして、ミズーリ工科大 (Sajal 教授) が持つ CPS 分野でのセキュリティフレームワークと CREST 山名チームで進める完全準同型暗号(FHE)の高速化手法を用い、電力スマートメータを事例にセキュアな CPS システムの構築を目標としており、大きな成果を期待している。既に、ICSDE2017(7/21-23)において共同研究のポジションペーパーを発表している (Sajal K. Das, Hayato Yamana: "Securing Big Data and IoT Networks in Smart Cyber-Physical Environments," Proc. of ACM International Conference on Smart Digital Environment 2017(ICSDE2017) (2017. 7. 21-23)。

NSFに加えEUとの連携を進めるため、2016年10月から仏国 INRIA の Director である Nozha Boujemaa 博士をアドバイザーとして招聘している。現在、仏国でのイベント開催を計画中である。

④「データさきがけ」とデータを介した連携

ビッグデータの研究を行うにあたり、利用可能なデータを確保することの重要性は論を俟たない。しかし、実際のビッグデータにアクセス可能な基盤技術の研究者が少ないことは、領域創設当初の段階で容易に想像できた。そのため、ビッグデータ応用領域とは異なり、実データの利用可能性を課題採択の必須条件とはしなかった。これは、応募条件を緩めることで間口を広げるメリットを重視した方針であり、実データがない場合のデメリットを、①から③で述べた連携の推進等によりある程度緩和可能と判断したものである。

一方、第1回目の募集では、特にさきがけで、様々なアプリケーション分野の研究者が多く応募してきた。アプリケーション分野に軸足を置いた研究者は、その分野の実データを利用できる可能性が高い。そこで、アプリケーション分野の研究者の力を活用し、領域全体の研究力向上を目指すため、特定のアプリケーション分野のデータを集取し、それを第三者が利用可能な形に整備することを目的とした「データさきがけ」を新設し、第2回目の募集から募集要項に記載した。「データさきがけ」については、第3回目の募集で3件の課題を採択している。「データさきがけ」によるデータの整理は順調に進んでおり、さきがけ3期生が整備したデータの一部は、既にCRESTチームなどに提供されている。

(3) 研究進捗状況の把握と評価、それに基づく指導

課題の進捗管理とアドバイスの提供は、次の2つの方法を組み合わせて行っている。

- (i) 領域内の研究者には公開された（一般向けには非公開の）領域会議
- (ii) 課題毎に分けて行う評価会とサイトビジット

各課題の進捗確認については、(i)の領域会議で年に1回は行っており、質疑の時間や終了後にはアドバイスも行う。研究分野の専門性が高いアドバイスは、関連分野に強いアドバ

イザーまたは正副の研究総括が行い、さらに、研究総括からは、(1)の運営方針で述べたような研究連携の強化やイノベーションに向けた取り組みに関するアドバイスも行う。例えば、(2)②であげたセキュリティとプライバシー保護を主要なテーマとする CREST 3 チームの研究連携はこのようなアドバイスに基づくものである。また、CREST 佐久間チームには 2015 年度の間評後、インパクトのある実証実験を促すアドバイスを行い、同時に研究費を増額することで、実証実験を加速させた。さらに、合同領域会議ではそのメリットを活かし、ビッグデータ応用領域の研究総括やアドバイザーからも応用の立場でのアドバイスを受けている。さきがけ課題については、さきがけ専用の領域会議においても、領域内での進捗確認とアドバイスを行っている。

領域会議は研究者間の情報共有にも有効だが、発表件数が多いため、1 課題あたりの時間はどうしても短くなる。そこで、より時間をかけた進捗確認は(ii)の方法で行っている。(ii)の頻度は1 課題あたり 1.5 年から 2 年に 1 回程度である。CREST 課題については、中間評価を通常の領域より 1 年前倒しで、課題開始後約 2 年目に行っている。個々の課題の進捗と今後の構想を確認し、チーム全体の今後の進め方に関するアドバイスを行っている。1 年前倒しにしているのは、アドバイスを反映するために十分な時間を確保するためである。さらに、その約 2 年後にサイトビジットを行い、中間評価のフォローアップ、終了に向けての取り組みに対するアドバイスなどを行っている。この段階では、いかにイノベーション創出につなげるかという観点でのアドバイスの比重を増やしている。

一方、さきがけでは、中間段階における時間をかけた進捗確認とアドバイスを、サイトビジットで行なっている。サイトビジットにおいては、個々の研究内容を対象としたものだけでなく、領域内での人的ネットワークの形成等の今後のキャリアを考えたアドバイスも行なっている。また、研究者の上長に対する協力依頼等も併せて行なっている。

(4) 研究費配分上の工夫

① 当初配分での工夫

ビッグデータ基盤技術の研究では、CREST・さきがけ共に、計算実験のピーク時に大量の計算資源を必要とし、それ以外のソフトウェア開発時等には少量の計算資源しか利用しないことが多い。このような場合、ピーク時に合わせて物理的な計算サーバを購入すると、平均利用率は低くなり、研究費の効率的な利用の妨げとなる。そこで、計算資源に関しては、可能な範囲でクラウド資源を利用するように募集要項に明記し、さらに各年度の予算承認時にも同様の指導を行っている。ただし、研究倫理審査でクラウド利用が認められない等の合理的理由がある場合には、もちろん、この方針を強制するものではない。

特に CREST においては、ビッグデータ関連の人材育成を進め、研究コミュニティを発展させるために、人への投資も重要である。物理計算サーバのようなモノよりも、若手研究者のような人や研究で用いるデータをより重視した配分を原則的に行なっており、CREST の総研

研究費の約半額は人件費・謝金に充当している。現在までに、研究員（エフォート 50%以上）55 名、RA 等の学生 43 名が雇用され、研究活動を通じた人材育成の対象となっている。なお、特殊な機器を必要とする研究では、モノへの配分も必要に応じて行なっている。

②研究の進展にともなう配分の変更

CREST の課題については、(1)の運営方針で述べた連携やイノベーション創出につながる研究を推進するために、配分額の変更を行なっている。研究の進展により、当初の計画段階では想定していなかった新しい可能性が芽生え、実証実験を行うことや成果のオープンソース化を進めることで将来の実用化が期待できる場合、特に重点的な研究費の増額対象とした。また NSF の研究者との連携促進のための費用も一部増額の対象とした。さらに、領域発足後に機械学習などの分野で急速な進歩が生じていることから、この潮流に対応しこれを利用することで、研究成果のレベルアップを図るための追加計画も一部増額の対象とした。

例えば、CREST 佐久間チームには前述のアドバイスに併せて、2015 年度～2017 年度に合計 3,240 万円を増額することにより実証試験を加速することができた。また、CREST 黒橋チームには、2016 年度と 2017 年度に合計 1,500 万円を増額することにより、対話エージェント構築の実証試験を協力自治体で行うことができた。

さきがけの課題については、アプリケーション分野の研究者が有するデータや知見を活用するための連携と「データさきがけ」で整備したデータを基盤技術の研究者に提供するための連携の促進のための費用を増額の対象とした。

(5)人材育成

AIP ネットワークラボへの貢献の一環として、CREST 課題に参加する優秀な若手研究者（大学院生を含む）を AIP チャレンジプログラムに推薦している。本研究領域から、2016 年度は 10 名、2017 年度は 13 名が採択され、AIP ネットワークラボから支給された研究費を用いて、個人研究に取り組んだ。2016 年度は、CREST の 3 研究領域から採択された計 29 名の若手研究者が一堂に会し、ポスター発表等で意見交換を行う場が設けられており、異分野連携に向けた訓練の場としても活用している（2017 年度も同様のイベントが予定されている）。なお、この 29 名の中から、本研究領域が推薦した胡緯華（CREST 原田チーム）がラボ長賞第 1 位の表彰を受けた。

本研究領域のさきがけ課題では、研究代表者自身が若手研究者に該当する。そのため、人的ネットワークの形成など、研究者としての今後のキャリアを考えたアドバイスも行っている。既にさきがけの研究期間を終了した 2013 年度採択の 1 期生 6 名については、それぞれ、さきがけ研究の成果を活かす形で新しい活動を始めている。具体的には、佐藤と田部井が理化学研究所 AIP センターのチームリーダーとユニットリーダーに就任、生貝が大学から民間企業に転身、松谷が研究代表者として他領域の CREST 研究を開始、水野と宮尾がさきが

けネットワークに採択され研究を開始している。なお、宮尾は 2015 年に日本学術振興会賞を受賞している。それぞれに今後のさらなる活躍が大いに期待できる。

6. 研究の経過と所見

(1) 研究総括のねらいに対する研究の状況

「2. 研究総括のねらい」で述べた内容に対し研究は順調に進んでおり、設定した 5 つのコア分野 (2(2)参照) のそれぞれで活発な研究活動が行われている。本研究領域の CREST 研究全体では、発表論文 911 編 (うち国際論文 850 編)、口頭発表 1,152 件 (うち招待講演 570 件)、特許出願 47 件 (うち国際出願 24 件)、さきがけ研究全体では、発表論文 93 編 (うち国際論文 68 編)、口頭発表 307 件 (うち招待講演 118 件)、特許出願 5 件 (すべて国内出願) の成果があがっている。このうち、コンピュータサイエンス分野で重視されるトップカンファレンスでの発表論文数は、CREST 研究から 72 編、さきがけ研究から 13 編に達している (「ERA で A ランク」と「Qualis で A1 ランク」を同時に満たす国際会議での発表論文のみを集計した)。また、ビッグデータに特化した国際会議として重要な IEEE Big Data での発表件数は、CREST 研究から 15 編、さきがけ研究から 1 編である。コンピュータサイエンス分野以外のアプリケーション分野の研究成果は、総合誌の Science、Nature Communications や各分野のインパクトファクターが高い雑誌にも掲載されている。特筆すべき研究成果については (3) (4) で述べる。

分野を超えた共同研究や NSF の研究者との国際共同研究も進みつつある。共同研究の件数は、本研究領域内が 6 件、ビッグデータ応用領域とのものが 3 件、その他の AIP ネットワークラボ所属領域とのものが 1 件、NSF の研究者とのものが 8 件、企業とのものが 11 件に達している。

(2) 研究領域全体として見た時の特筆すべき成果の見通し

本領域の運営にあたって見据えているのは、将来の科学・産業・社会を支えるデータ基盤・サービスプラットフォームであり、その実現のための鍵となる技術こそが特筆に値する。このような観点から、本領域の今までの研究成果を俯瞰すると、次のような成果があがっている。

- ビッグデータのモデリングや解析手法の適用範囲の拡大と桁違いの高速化
- 自然言語や画像のデータからの意味抽出の高度化
- 従来のものよりはるかに大規模なデータ処理を実現するシステムの設計
- 解析の全過程を暗号化したまま行う方式の実用レベルの性能達成

より詳しくは (3) (4) で述べるが、前半の 2 項目はプラットフォームが提供する API や機能の強化を担い、後半の 2 項目は実行系の高性能化とデータの強靭化を担うものである。

(3) 科学技術の進歩に資する研究成果

①CREST 研究の主要な成果

CREST 研究では、4つのコア技術分野で多くの学術的に優れた研究成果が得られ、論文発表等を行っている。各分野の代表的な成果を以下で簡単に説明する。なお、ビッグデータ法制度分野の CREST 研究は、暗号技術を用いた個人情報保護に関する政策提言等を行い、より優れた技術の普及を促すことを主に目指しており、直接的に「科学技術の進歩」を目指すものではないが、パブリックコメントに対する意見表明等の活動を行っている。

数理基礎技術分野では、CREST 山西チームが、観測データに直接は現れない系の潜在的な構造を抽出するために有効な普遍性の高いモデル選択規準を確立した。この選択規準は適用範囲が広く、従来の統計的モデル選択規準の適用が難しかった潜在変数モデルの広いクラスに対しても、高い計算性能で適用が可能である。潜在変数モデルは機械学習で有用なものであり、この研究のインパクトは非常に大きいことが期待できる。この成果は KDD、ICDM などのデータマイニング分野のトップカンファレンスで発表されている。また、CREST 原田チームの杉山グループは、効果的な機械学習のためには大量のラベル付き教師データが必要になるという問題への取り組みで、大きな成果をあげた。大量のラベル付き教師データがなくても正例とラベルなしデータだけから高い収束率で学習が可能なる手法の開発とその理論的背景を明らかにした。これにより、ラベル付けの手間を大きく削減したり、大量の教師付きデータが存在しない分野での機械学習の利用を促進したりすることが期待できる。一連の論文が NIPS や ICML などの機械学習分野のトップカンファレンスで発表されている。これは世界に先駆けた成果である。その後、これらのトップカンファレンスで、フォロワーの発表も行われるようになっており、科学技術の進歩に大きく貢献している。

自然言語・画像基盤技術分野では、本領域発足後に急速に進歩したニューラルネットを用いた言語処理において、CREST 黒橋チームが、日本語の形態素解析、構文解析、照応省略解析などで先駆的な研究を行い、ACL や EMNLP などの自然言語処理分野のトップカンファレンスで発表を行なっている。日本語の構文解析と照応省略解析については、世界最高性能の解析精度を実現している。我が国の知識基盤を充実させるためには、日本語の解析技術の高度化が必須であり、これは日本語を用いて知的作業を行う者にとって重要な意義を持つ可能性がある研究成果である。

セキュリティ基盤技術分野では、従来、実用化のためには数桁の高速化が必要と考えられていた完全準同型暗号（データを暗号化したまま演算を行うことが可能な暗号方式）について、CREST 佐久間チームがブレークスルーを生み出した。具体的には、実用的な規模のゲノムワイド関連解析（20 万 SNP、1 万サンプル）を、実用的な計算時間（8 core 並列で 8 時間）とメモリ量（10MB 程度）で達成可能なことを示した。これはマイクロソフト社による本研究以前の世界トップクラスの結果と比べ、計算時間で 1000 倍近くの高速度化、メモリ量で 2500 倍近くの高効率化を達成したものである。プライバシー情報等の漏洩リスクが従来のものより

きわめて低い方式を実用化に大きく近づけるものであり、顕著な成果である。

システムアーキテクチャ基盤技術分野では、CREST 松岡チームが、きわめて大規模なビッグデータ処理のワークロードの分析とその処理に適したアルゴリズムやミドルウェアの開発およびオープンソース化を行った。ビッグデータとして多くの応用が期待される巨大グラフの高速処理に関しては、CREST「ポストペタスケール高性能計算に資するシステムソフトウェア技術の創出」の藤澤チームと共同で、国際的なベンチマークである Graph500 において、2015 年 6 月から 6 回連続（年に 2 回）で世界一の成果をあげている。そして、これらの知見を、ビッグデータと HPC(High Performance Computing)の融合を果たした世界トップレベルのスーパーコンピュータの設計にも活用した。この研究の成果は、広範な科学技術分野で利用可能な強力なツールを提供するものであり、科学技術の進歩に大きく貢献するものである。

②さきがけ研究の主要な成果

さきがけ研究では、セキュリティ基盤技術分野以外の 3 つのコア技術分野とビッグデータ法制度分野で研究を行っている。また、本研究領域の特徴的活動であるデータさきがけでも研究が進んでいる。代表的な成果を以下で簡単に説明する。

数理基礎技術分野では、1 期生の佐藤が、統計的潜在意味解析を可能とする代表的なモデルである LDA(Latent Dirichlet Allocation)の大規模なデータからの学習において、従来手法では問題となっていたロングテールを含むデータの高速処理を可能とする方式を考案した。この結果に至る一連の成果は、トップカンファレンスの KDD、ICML で発表している。佐藤はこの他に、東京大学医学部附属病院と協力し、機械学習の読影への適用に関する研究も行った。既にさきがけの研究期間は終了しているが、この研究協力関係は継続しており、今後の発展が期待できる。同じく 1 期生の田部井は、データを圧縮したまま解析を行う方式に関する一連の研究を行い、トップカンファレンスの KDD や ISMB で発表している。田部井の研究成果のうち大規模化合物データベースの類似度検索技術に関するものは、ビッグデータ応用の船津チームの研究にも貢献しており、他にもさまざまな応用の可能性がある。両者とも基盤技術の研究で優れた成果をあげ、その結果を応用に結び付けており、将来の発展が期待できる。

自然言語・画像基盤技術分野では、1 期生の宮尾が、自然言語のテキスト間をつなぐ含意関係認識、自然言語のテキストと記号表現された非自然言語のデータをつなぐ質問文のデータベースクエリへの変換、自然言語のテキストと非記号表現の画像をつなぐ画像検索とそのための意味表現に関する研究等を行った。一連の成果は、自然言語分野のトップカンファレンスである ACL、COLING、EMNLP など発表している。全体として見ると、自然言語、画像、データベースという異なるタイプのビッグデータを、統合的に活用するための基盤技術に関する研究であり、将来大きなインパクトをもたらす可能性がある。

システムアーキテクチャ基盤技術分野では、1 期生の松谷が、4 種類の NoSQL (キーバリ

ューストア型、カラム型、ドキュメント指向型、グラフ型)、ストリーム処理、バッチ処理などの特性の異なる処理に対し、既存のソフトウェアスタックを前提に、FPGA や GPU をアクセラレータとして活用し、省電力化と高性能化を目指す研究を行った。一連の研究が国際会議等で発表されており、HEART(International Symposium on Highly-Efficient Accelerators and Reconfigurable Technologies)において Best Paper Award を受賞した。リアルタイム処理とバッチ処理を組み合わせ、省電力かつ高性能に、さまざまなタイプのビッグデータ処理を行うシステムアーキテクチャの主要構成要素を網羅した研究であり、将来の情報システムに大きな影響を与える可能性がある。

ビッグデータ法制度分野では、1 期生の生貝が、個人情報保護やオープンデータに関連した法制度の日米欧比較等に関する研究を行った。これらは、パーソナルデータや知的財産権の保護とデータの有効活用のバランスについて熟慮し、社会的合意を取る必要があるテーマである。単に欧米の制度を輸入するだけでは済まず、文化的背景の違いも理解して、我が国に適した制度設計を行うことが重要である。比較研究の結果を整理し、主に国内向けに成果の公開を行っており、新聞への寄稿などの一般向けの情報発信も行っている。技術の社会受容に関わる問題であることから、一般向けの啓発活動にも重要な意義がある。

データさきがけでは、3 期生の小野木、柳澤、山田がデータの収集と整理を進めている。小野木は、全国規模で 50 年間以上にわたる大豆の栽培データの収集・整理を行っている。柳澤は、頭蓋内の電極による脳活動データとそれに対応した行動データや視聴覚データの収集・整理を行っている。山田は人の腸内環境データとそれに対応した臨床データや生活データの収集・整理を行っている。このようなデータは利用可能な形で今まで存在しなかったものであり、関連する科学技術の進歩に貢献することが期待できる。

(4) 科学技術イノベーションに貢献する成果

数理基礎技術分野では、CREST 山西チームが、機械学習やデータマイニングなどの新しいアルゴリズムを実世界の問題に適用し、実用化を目指す試みを行っている。代表的なものとしては、緑内障進行予測に関する研究で、従来の緑内障の診断方式よりはるかに被検査者の負担が軽い方式の開発が進んでいる。これは、OCT(Optical Coherence Tomography)を用いて網膜の厚みを測り、その結果から診断を行う世界初の試みである。緑内障は失明に至りうる病であり、社会の高齢化に伴い、今後ますます深刻化が予想される問題への取り組みとして重要なものである。この研究は、東京大学医学部附属病院眼科の協力を得て行っている。

自然言語・画像基盤技術分野では、CREST 黒橋チームにより、企業や自治体と協力した取り組みが行われている。一例をあげると、民間企業が集めた 700 万件超の生活者の意見(製品に対する不満等)を分析し、それを製品の改善等に利用するために、本領域の研究成果が活用されている。また、別の事例として、民間企業および自治体と協力し、自治体のホームページ中のテキストを自動解析してサービスに関する問い合わせに回答する対話エージェントの構築などを行いつつある。これらは、日本語で記述されたデータベース化されていない

い散在情報を、集約化し有効に活用するための取り組み事例である。人手での活用が困難な散在情報は、これらに限らず我々の社会の様々な場所に散らばっていることが予想される。その活用を進めることには、大きな社会的意義がある。

セキュリティ基盤技術分野では、CREST 佐久間チームが、個人のゲノム情報と医療情報を対象に、収集・統合・統計解析の一連の作業を暗号化したまま実施する秘密計算システムの開発を行い、病院や民間企業と協力して、生活習慣病の予防医療を支援する実証実験を行っている。秘密計算システムを用い、複数の医療施設からの情報を統合して医療に応用する実証実験としては世界初のものである。複数の医療施設のデータを統合することで、単独の医療施設のデータのみからでは抽出困難な情報の分析ができることは明らかであり、高度な機密保持と統合という相矛盾する要件を満たすこの実証実験には、大きな意義がある。

システムアーキテクチャ基盤技術分野では、東京工業大学の TSUBAME3.0 や産業技術総合研究所の ABCI (AI Bridging Cloud Infrastructure) などの AI 性能の高いスーパーコンピュータの設計に、CREST 松岡チームの研究から得られた知見が活用されている。これらは、学術応用から産業応用まで広範な利用が期待でき、イノベーション創出の促進につながる事が期待できる。

7. 総合所見

当初のねらいに対して、研究は順調に進んでいる。5つのコア分野を設定し、集中と分散のバランスを取りながら課題を採択したことで、「6. 研究の経過と所見」で述べたように、学術的に優れた成果、将来のイノベーション創出につながる可能性のある成果が生まれている。最も基本的な研究活動として、今後も、各コア分野の研究推進をはかるべきである。領域としてはこれから終盤に近づくため、各研究者の意識を、今まで以上にイノベーション創出に向けることも重要と考えている。

さきがけと CREST の複合領域としては、両者の特性の違いを考慮し、本研究領域全体の研究力向上につながるような課題選考と領域運営を行っている。さきがけ研究者には、尖った基盤技術またはアプリケーション分野の知見とデータに根ざした挑戦的な研究を期待している。一方、CREST 研究者には、より大きなスケールとイノベーション創出への展開を期待している。研究領域全体としては、CREST 課題でビッグデータの中核的なテーマに確実に取り組みつつ、さきがけ課題で、個人研究ゆえに軽やかに動けるメリットを活かした挑戦や「データさきがけ」のように CREST でカバーできない活動を行っている。専門とする分野のみならず研究スタイルの多様性を高めた上で、課題間の連携を促進し、シナジーの創出を目指す運営を行っている。

将来の見通しに関しては、各コア技術分野の研究活動から生まれる基盤技術を集約し、データ基盤・サービスプラットフォームの構築につなげることが重要である。「6. 研究の経過と所見」で述べたように、将来のデータ基盤・サービスプラットフォームの基盤技術となりうるものが、少なくとも萌芽的な形では現れつつある。本研究領域の役割を考えると、これ

らを発展させることに注力するのが妥当と思われる。データ基盤・サービスプラットフォームの構築に本研究領域の成果を活かすことで、社会的意義が認められるレベルに達すると予想している。

領域内外の連携の推進についても、「5. 研究領域の運営について」で述べたように、運営上の重要事項として取組み、領域内連携、ビッグデータ応用領域との連携、NSF のビッグデータ関連プロジェクトとの連携のそれぞれで、共同研究の事例が生まれている。しかし、分野の壁や文化の壁を超えるには、時間と労力がかかる。そのため、急峻な立ち上げを期待すべきではない。特に、NSF との連携は昨年度から始めたばかりであり、始める前には予期し得ないような全く新しいシナジーが生まれるまでには、もう少し時間がかかると考えている。

このような連携を進めることで、本領域の終了後にも良い影響を残すことが期待できる。繰り返しになるが、分野の壁や文化の壁を研究者が超えるためには多大な労力を要する。しかし、1人で挑戦するより、経験者のサポートを得て挑戦する方がはるかに楽である。本領域の活動により、連携の経験がある人材が育ち、その人材がハブやブリッジとして研究者同士をつなぐ構造ができあがると、さらに大きなシナジーが生まれる可能性がある。

人材以外の本研究領域の成果として、論文も重要であるが、生み出されたデータやソフトウェアも重要である。従来は、研究者の評価が主に論文により行われたため、データやソフトウェアは論文を書くために必要な範囲で整備する傾向が強かった。しかし、そのような考え方では、他の研究者等が円滑に利用できるデータやソフトウェアは生まれにくい。研究者コミュニティ全体にとっては、他の研究者等が円滑に利用できるようにすることに大きな価値がある。しかし、データやソフトウェアをそのレベルまで整備するには、より多くの労力が必要になる。一度出版されたら通常は改訂の必要がない論文とは異なり、データやソフトウェアを使い続けるために必要なメンテナンスのコストも無視できない。一方で、人間にとっての知識源である論文と機械にとっての知識源であるデータやソフトウェアの重要度を比べたとき、時間とともに次第に後者に傾いて行くのは必然である。ビッグデータ研究の推進は、この傾向を加速させるものでもある。このような時代の変化を反映した研究施策の実施が望まれる。

以上