

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」

複合領域事後評価用資料

研究総括：喜連川 優

副研究総括：柴山 悦哉

2021年2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)採択研究課題・研究費	7
2. 研究領域および研究総括の設定について	10
(1)研究領域設定の理由	11
(2)研究総括指定の理由	12
3. 研究総括のねらい	13
(1)重点的に取り組むべきコア分野の設定	13
(2)研究用データの整備	14
(3)分野を超えた連携	15
(4)国際連携	15
4. 研究課題の選考について	15
(1)選考方針	15
(2)CRESTの選考過程と選考結果	16
(3)さきがけの選考過程と選考結果	17
5. 領域アドバイザーについて	18
(1)人選にあたっての考え方	18
(2)CREST	18
(3)さきがけ	20
6. 研究領域のマネジメントについて	22
(1)研究総括の運営方針	22
(2)領域内外との連携・協力の推進	22
(3)研究進捗状況の把握と評価、それに基づく指導	26
(4)研究費配分上の工夫	27
(5)複合領域中間評価結果への対応	28
(6)人材育成	30
7. 研究領域として戦略目標の達成状況について研究を実施した結果と所見	31
(1)研究総括のねらいに対する研究成果の達成度	31
(2)研究領域全体として見た時の特筆すべき研究成果	31
(3)科学的・技術的な観点からの貢献	32
(4)社会的・経済的な観点からの貢献	37
(5)本研究領域に続く研究資金の獲得状況	39
8. 総合所見	40
(1)研究領域のマネジメント	40
(2)戦略目標の達成状況	40

(3) 本研究領域を設定したことの意義.....	41
(4) 科学技術イノベーション創出に向けた、今後への期待、展望、課題.....	41
(5) その他の所感	42

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、強化するとしている。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-X、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）

本研究領域は、戦略目標の2つの達成目標（1(1)①(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・役職は領域終了時点)

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) 採択研究課題・研究費

①CREST

(百万円)

採択年度	研究者	所属・役職 上段：研究終了時 (下段：採択時)	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	425
	佐久間淳	筑波大学 教授 (筑波大学 准教授)	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 特任教授 (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	380
	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	337
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	343
	加藤直樹	兵庫県立大学 学部長・教授 (京都大学 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	324
	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	415
	宮地充子	大阪大学 教授 (北陸先端科学技術大学院大学 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	337
	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	307
	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	297

			キュアなコンテンツ共有・流通基盤の構築	
			総研究費	3,830

各研究課題とも研究期間の総額、進行中の課題は予定を含む(2020年10月31日現在)

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」(6(4)参照)という考え方に基づいた配分を行った。総研究費の半分近くが人件費・謝金に充当されており、これに旅費を加えると約2/3になる。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、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、米国 NSF を含む領域内外の研究者との連携の推進、基盤技術と応用分野の橋渡しを行う人材の雇用、アウトリーチ活動に対し、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さきがけ

(百万円)

採択年度	研究代表者	所属・役職 上段：研究終了時 (下段：採択時)	研究課題	研究費*
2013 年度	生貝直人	東京大学 客員准教授 (新領域融合研究センター 特任研究員)	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	9
	佐藤一誠	東京大学 講師 (東京大学 助教)	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	38
	田部井靖生	東京工業大学 さきがけ専任研究者	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	20
	松谷宏紀	慶應義塾大学 講師	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	金融ビックデータによるバブルの早期警戒技術の創出	42
	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
	杉山鷹人	国立情報学研究所 准教授 (大阪大学 助教)	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ング法の開発	28
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	8

	小野木章雄	農業・食品産業技術総合研究機構 主任研究員(東京大学 農学研究員)	膨大なレガシー栽培データを蘇生する(データさきがけ)	27
	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	横浜市立大学 特任教授 (京都大学 助教)	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	40
	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 教授 (大阪大学 助教)	思考・行動を予想する脳ビッグデータ (データさきがけ)	33
	山田拓司	東京工業大学 准教授 (東京工業大学 講師)	ヒト腸内環境ビッグデータ (データさきがけ)	36
			総研究費	563

*各研究課題とも研究期間の総額

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及び米国 NSF 研究者との連携の推進に対し、重点的に増額を行った。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。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、そして NSF 研究者との連携を推進するため、4名に計 10.4 百万円の重点的な増額を行った。

2. 研究領域および研究総括の設定について

本戦略目標の下、相互に関連する以下の二つの研究領域が設定された。

研究領域 1 「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」(CREST)

研究領域 2 「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」
(CREST・さきがけ複合領域)

研究領域 2 が本研究領域である。

(1) 研究領域設定の理由

本戦略目標は、分野を超えたビッグデータの利活用により、新しい科学的発見や社会的課題解決に資する新たな知識や洞察を得るための革新的な情報技術や数理的手法の創出・高度化・体系化を目指すものである。このためには、生命・健康・医療から地球環境、都市基盤システム、防災・減災までの多様な分野における科学的発見・社会的課題解決を目指す横断型アプローチと、ビッグデータを利活用する上で本質的となる技術課題を情報学・数理工学的見地から深く掘り下げブレークスルーをねらう深堀型アプローチによる重奏的な研究が行われる必要がある。そのため、本戦略目標の下には、2つの研究領域を選定することが適切である。

とくに、研究領域 1 における次世代アプリケーション技術の創出・高度化に当たっては、研究領域 2 の研究で得られる IT 基盤技術を取り込みながら推進し、その逆に研究領域 2 の IT 基盤技術の創出・高度化・体系化に当たっては、研究領域 1 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、研究領域 1 と研究領域 2 が相互連携することで、効果的に戦略目標が達成されるものと考えられる。

研究領域 1 においては、科学的発見及び社会的・経済的な挑戦的課題の解決や革新的価値創造のために、個々の研究者や研究室のみでは集積することが困難な大規模かつ多様な関連データを分野や組織を越えて集積し、相互に関連付けて高度な統合的分析処理を行うことにより、これらのビッグデータからそこに隠されている革新的新知見や価値を抽出し創成することを実証的に研究開発することを目的とする。さらに、この目的の達成に必要な IT 基盤技術を実証的に創出・高度化・体系化することを目指す。

そのためには、生命、物質材料、健康・医療、社会・経済、都市基盤システム、防災・減災、農林水産業、宇宙地球環境など様々な IT 以外の研究者と IT 系研究者との連携が不可欠であるが、現状ではデータの活用についてはまだ端緒の段階にあり、このような連携はあまり多く見られない。本研究領域の設定により、このような連携による研究を促すことで、優れた研究提案が数多く見込まれるとともに幅広く数多くの研究分野の発展に寄与することが期待される。

研究推進体制としては、上記のように IT 系以外と IT 系の研究者がチームを組むことができる CREST を選定することが適切である。

以上のことから、本研究領域は戦略目標の達成に向けて適切に設定されている。

研究領域 2 においては、ビッグデータの高度な統合利活用により、新しい科学的発見による知的価値の創造や、それらの知識の発展による社会的・経済的価値の創造やサービスの

向上・最適化などに資する、情報学を基盤とした IT 基盤技術の創出・高度化・体系化を目指す。このため、大規模データを圧縮・転送・保管する大規模管理システムの安定的運用技術や、多種多様な情報を横断して検索・比較・可視化して真に必要な知識を効率的に取り出す技術、これらを可能にする数理的手法やアルゴリズムなどの開発を対象としている。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全体のメカニズムデザインやエコシステムを視野に入れ、個別要素技術を組み合わせ総合的に高度化・体系化する取り組みが必要とされている。そのため、関係する研究者がチームを組んで研究推進する CREST を選定することが適切である。

また、独創的なアプローチを有するポテンシャルの高い個人研究者について当該研究領域への参画を促すことで、ビッグデータ時代に必要となる IT 基盤技術を研究開発する研究者コミュニティが醸成されると見込まれる。

さらに、ビッグデータのドメインに依存しない本質的課題を抽出・解決する上で、個人研究者が研究チーム等の研究目標に捉われず、自由に IT 以外の分野との横断的交流を進めることは、将来的な IT 以外の分野との積極的な連携・融合に資すると期待される。したがって、さきがけを選定することは適切である。

本研究領域の設定により、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・体系化への取り組みが行われるとともに各アプリケーション分野における次世代アプリケーション基盤技術の創出・高度化へ資することから、本研究領域は戦略目標の達成に向けて適切に設定されている。また、ビッグデータから新たな知識等を得ようとする社会的ニーズが高まる中、そのための技術・手法の創出・高度化・体系化を促すものとして適切な研究領域であり、データマイニング・機械学習・統計学等の分野から優れた研究提案が多数見込まれる。

(2) 研究総括指定の理由

喜連川優氏は、web マイニングを用いたサイバーコミュニティ抽出と次世代サーチエンジンの研究、大規模 PC クラスタによる超並列データマイニングおよびデータベース研究など、多数の卓越した業績をもつ研究者である。世界に先駆けてビッグデータの重要性に着目し、「情報爆発時代に向けた新しい IT 基盤技術の研究」(科学研究費補助金特定領域研究、平成 17～22 年度) や、「超巨大データベース時代に向けた最高速データベースエンジンの開発と当該エンジンを核とする戦略的社会サービスの実証・評価」(最先端研究開発支援プログラム (FIRST)、平成 22～26 年度) のプロジェクト代表者として、ビッグデータから真に必要な情報を効率よく取り出す要素基盤技術や、社会が受け入れやすい先進的 IT サービスのためのシステム構築に大きく寄与している。優れた研究成果は、電子情報通信学会業績賞、情報処理学会功績賞の受賞を始め、リレーショナル・データベースの結合演算の高速化への貢献により ACM SIGMOD Edgar F. Codd Innovations Award を日本人として初めて受賞するなど、国内外で高く評価されている。これらのことから、同氏は、ビッグデータ時

代における IT 基盤技術について深い先見性と洞察力を有すると判断される。

また、国立国会図書館科学技術関係資料整備審議会委員や学術会議会員などを歴任し、かつ IEEE ICDE や PAKDD など複数の国際会議のチェアを務めるなど、当該分野を牽引する世界的リーダーの一人である。その発展への精力的な関与により、多数の分野における豊富な人脈を有し、国内外の関連分野の研究者からの信頼も厚く、適切な研究マネジメントを行う経験、能力を有するとともに、公平な評価を行いうると認められる。

加えて、東京大学地球観測データ統融合連携研究機構長として、地球環境分野と IT 分野の連携・融合に積極的に取り組むなど、IT 以外の分野との融合を目指す本戦略目標の達成に不可欠となる、俯瞰的視野を有すると見られる。

以上より、同氏は本研究領域の研究総括として適任であると判断される。

3. 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な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、第三者の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

データの整備はあまり進んでいないのが実情である。そこで、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

本研究領域は、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研究領域と 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、本研究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。

4. 研究課題の選考について

(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)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。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。また、チーム内のシナジーを重視し、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分野

(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・画像基盤技術分野、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。

(3) 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、CREST の選考と同様に、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。

5. 領域アドバイザーについて

(1) 人選にあたっての考え方

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「クラウドコンピューティング」、「OS」、「セキュリティ」、「画像処理」、「人工知能」、「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」、「ユビキタスコンピューティング」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（1(1)①(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た。なお、領域アドバイザーのうち3名が、途中から別のCREST及びさきがけ領域の研究総括に就任し、他領域においても活躍している。

(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 (専門分野)	終了時の所属 (着任時の所属)	役職 (着任時の役職)	任期
荒川 薫 (画像処理)	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

(人工知能)			
岩野 和生 ※ (クラウドコンピューティング、サイバー・フィジカルシステム)	東京工業大学 (三菱商事株式会社)	客員教授 (顧問)	2013年6月～2015年10月
上田 修功 (機械学習)	NTT コミュニケーション科学基礎研究所	フェロー	2013年6月～2021年3月
柴山 悦哉 ※ (プログラミング言語、セキュリティ)	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 (コンピュータアーキテクチャ、セキュリティ)	岩崎学園 (情報セキュリティ大学院大学)	理事 (学長)	2013年6月～2021年3月
辻井 潤一 (自然言語処理)	産業技術総合研究所 (マイクロソフト・リサーチ・アジア)	研究センター長 (首席研究員)	2013年6月～2021年3月
徳田 英幸 (OS、ユビキタスコンピューティング)	情報通信研究機構 (慶應義塾大学)	理事長 (教授)	2013年6月～2021年3月
徳山 豪 (アルゴリズム)	関西学院大学 (東北大学)	教授 (教授)	2013年6月～2021年3月
東野 輝夫 (モバイル・ユビキタスコンピューティング)	大阪大学	教授	2013年6月～2021年3月
堀 浩一 (人工知能)	東京大学	教授	2017年12月～2021年3月
室田 一雄 ※ (数理最適化)	東京大学	教授	2013年6月～2015年9月
安浦 寛人 ※	九州大学	理事	2013年6月～2014年3月

(コンピュータアーキテクチャ)		副学長	
Calton Pu (クラウドコンピューティング)	Georgia Institute of Technology (米)	Professor	2013年6月～2021年3月
Nozha Boujema (マルチメディア検索、人工知能)	Median Technologies (仏) (Inria (仏))	Chief Science & Innovation Officer (Director of Research)	2016年10月～2021年3月

(3) 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 (専門分野)	終了時の所属 (着任時の所属)	役職 (着任時の役職)	任期
荒川 薫 (画像処理)	明治大学	学部長 教授	2013年6月～2019年3月
石塚 満 (人工知能)	東京大学	名誉教授	2013年6月～2019年3月
岩野 和生 ※ (クラウドコンピューティング、サイバー・フィジカルシステム)	東京工業大学 (三菱商事株式会社)	客員教授 (顧問)	2013年6月～2015年10月
上田 修功 (機械学習)	NTT コミュニケーション科学基礎研究所	フェロー	2013年6月～2019年3月
柴山 悦哉 ※ (プログラミング言語、セキュリティ)	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 (コンピュータアーキテクチャ、セキュリティ)	岩崎学園 (情報セキュリティ大学院)	理事 (学長)	2013年6月～2019年3月

ティ)	大学)		
辻井 潤一 (自然言語処理)	産業技術総合 研究所 (マイクロソ フト・リサー チ・アジア)	研究センター長 (首席研究員)	2013年6月～2019年3月
徳田 英幸 (OS、ユビキタスコン ピューティング)	情報通信研究 機構 (慶應義塾大 学)	理事長 (教授)	2013年6月～2019年3月
徳山 豪 (アルゴリズム)	関西学院大学 (東北大学)	教授 (教授)	2013年6月～2019年3月
東野 輝夫 (モバイル・ユビキタ スコンピューティン グ)	大阪大学	教授	2013年6月～2019年3月
堀 浩一 (人工知能)	東京大学	教授	2017年12月～2019年3月
室田 一雄 ※ (数理最適化)	東京大学	教授	2013年6月～2015年9月
安浦 寛人 ※ (コンピュータアー キテクチャ)	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) (データベース)	筑波大学	教授	2013年6月～2019年3月
山西 健司 (さきがけ専任) (データマイニン グ)	東京大学	教授	2013年6月～2019年3月
Calton Pu (クラウドコンピュ ーティング)	Georgia Institute of Technology (米)	Professor	2013年6月～2019年3月
Nozha Boujemaa	Median	Chief Science &	2016年10月～2019年3月

(マルチメディア検索、人工知能)	Technologies (仏) (Inria (仏))	Innovation Officer (Director of Research)	
------------------	------------------------------------	--	--

6. 研究領域のマネジメントについて

(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、「3. 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。本研究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た。

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た。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うこととした。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た。研究課題の進捗状況の把握と評価、それに基づく指導については、(3)で詳しく述べる。

「3. 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た。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回開催した。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た。領域内外との連携・協力の推進については、(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た。

(2) 領域内外との連携・協力の推進

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た。このような試みから、基盤技術を有する本研究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た。代表的なものとしては、本研究領域さきがけ 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行うなど、積極的に他分野連携を進めた。

さきがけ研究者については、積極的な連携活動が、個人研究からチーム型研究への発展につながるケースがみられた。1 期生の田部井は、他領域のさきがけ研究者が代表者を務めた AIP 加速 PRISM 研究（官民研究開発投資拡大プログラム（PRISM）推進費を活用し、AIP ネットワークラボにおける特定成果の強化・加速をはかる研究）で研究分担者として活躍した。また、さきがけ研究の成果をもとに、AIP ネットワークラボや他領域のチーム型研究に採択された事例が 4 件ある。1 期生の松谷と 3 期生の柳澤は CREST 課題、3 期生の島田と山田は AIP 加速課題（AIP ネットワークラボの領域における優れた研究成果をベースに、新たな方向付けをした研究課題）に採択され、さきがけ研究の後継となる研究をそれぞれ代表者として進めている。

②領域内連携

本研究領域内部での連携についても、次のような点を考慮して、推進を図った。

- (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行なった。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とした。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た。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行われた。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い、「さきがけネットワーク」に採択された。また、後者の代表的事例としては、さきがけ 3 期生の柳澤が「データさきがけ」で収集・整備した脳情報のデータを CREST 原田チームの技術を用いて解析した。また、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た。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み、2017 年と 2018 年には、3 チームの協働による国際ワークショップ (IEEE International Workshop on Big Data and IoT Security in Smart Computing) を開催した。佐久間チームの課題は 2018 年度終了のため、2019 年度以降は「3 チームの協働」とは呼べないが、2019 年と 2020 年にも同ワークショップは継続して開催された。なお、このワークショップは、次の③で紹介する国際連携の成果でもある。また、システムアーキテクチャに関する研究を進めた 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た。

③国際連携

米国 NSF や EU 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た。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た。

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ない、2019 年度までに日本で 5 回、米国で 4 回開催した (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た)。2020 年度はコロナ禍のため、合同領域会議と公開シンポジウムをオンラインで開催とした。2020 年度の合同領域会議では、一部の時間帯で NSF 研究者に米国から遠隔参加頂いた。また、同年度の公開シンポジウムでは予め録画したビデオで講演して頂いた。

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た。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った。また、運営に関する意見交換も行い、NSF 側でも、Dear Colleague 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いた。

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 回開催した。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た。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た。本研究領域と NSF の研究者間での共同研究の代表的事例としては、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)) との共同研究がある。JST-NSF joint Big Data Workshop を契機として 2016 年に交流が始まり、2017 年以降は「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 システムの構築に関する研究を実施した。②で紹介した国際ワークショップ (IEEE International Workshop on Big Data and IoT Security in Smart Computing) の開催には Sajal 教授も協力している。また、NSF と NICT が共同で公募した日米連携事業にも両者の共同研究課題が採択された。

米国に加え EU との連携を進めるため、2016 年 10 月に、当時仏国 INRIA の Director であった Nozha Boujemaa 博士をアドバイザーとして招聘した。Boujemaa 博士には課題の事後評価等で助言を頂くとともに、仏国 DATAIA とのシンポジウムの合同開催でもご尽力を頂いた。その結果、2018 年 7 月にはパリで DATAIA との共同でシンポジウムを開催し、2019 年度には日本でシンポジウムを 2 回 JST、NSF、DATAIA の共同で開催した。これらの活動により、EU の研究者との交流を図った。なお、2020 年度にも EU において合同イベントを開催することを計画していたが、コロナ禍のため、この計画は断念せざるをえなかった。

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、「データさきがけ」を新設した。これは、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さきがけ課題であり、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 件の課題を採択した。これら 3 件の課題では、ヒトに関する機微なデータや農業の競争力に関わるデータの収集・整備を行なったため、完全なオープン化はできなかったが、一定の条件を満たす研究者へのデータ提供が行われている。確認できている範

囲で、現在までに 16 編の論文がこれらのデータを利用して公表されている。うち 8 編は、20 程度以上のインパクトファクターを有する論文誌に掲載された（内訳は、Nature Medicine に 4 編、Gut に 2 編、Circulation に 2 編）。

「データさがけ」で収集・整備したデータを、さがけ研究終了後に、いかに維持して継続利用可能とするかが問題になることは、「データさがけ」を考案した当初から認識していた。結果的には、柳澤と山田が CREST 課題と AIP 加速課題にそれぞれ採択され、小野木が農研機構に異動したことで、単に集めたデータを維持するだけでなく、さらなる収集・整備を進めることまで可能となった。本研究領域で蒔いた種が、大きく育つ可能性につながった。

(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で行った。

- (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 回は行い、質疑の時間や終了後にはアドバイスも行った。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った。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 年度の間評価の後、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した。その結果、複数の医療施設からの情報を統合して医療に応用するものとしては、当該分野で世界初の実証実験を行うことができた。

領域会議を合同で行うメリットを活かし、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た。また、NSF 研究者を合同領域会議等に招待していることで、アドバイスの有効性が増したケースもあった。一例をあげると、合田チームに対しては、複数のクラウドを結ぶ実証実験を行う際に、国内に閉じず、グローバルな学術基盤形成を意識するように助言した。その際に、NSF が支援するクラウドサービスとの連携の可能についても助言した。その結果、Chameleon Cloud との連携が実現され、日米にまたがる仮想ネットワークをオンデマンドで作成する仕組みを構築することができた。さがけ課題については、さ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た。

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)の方法で行った。(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 年目に行った。個々の課題の進捗と残り約

3.5年間の構想を確認し、チーム全体の進め方に関するアドバイスをを行った。1年前倒しにしたのは、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た。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た。

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た。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た。また、研究者の上長に対する協力依頼等も併せて行なった。

(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た。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、この方針の例外とした。

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、若手研究者等の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た。CRESTの総研究費の約46%は人件費・謝金に充当された。研究員（エフォート50%以上）67名、RA等の学生167名が雇用され、研究活動を通じた人材育成の対象となった。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た。

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た。研究の進展や環境の変化により、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。また米国NSFの研究者等との海外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。その他に、研究成果であるビッグデータ基盤技術を応用分野で活

用するための橋渡し人材の雇用やアウトリーチ活動も増額の対象とした。

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～2018 年度に合計 1,800 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。2017 年度には、米国 NSF の研究者とのマッチングのためにアトランタでワークショップを開催し、その結果始まった共同研究を支援するために、3 チームの 8 課題に対して計 1,520 万円の増額支援を行なった。2020 年度には、加藤チームから提案があったコロナ禍に対応した 2 件のテーマに対し、計 970 万円の増額を行なった。

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) 複合領域中間評価結果への対応

① 高く評価いただいた施策の継続

「データさきがけ」、海外連携など、本研究領域のマネジメントについて領域中間評価で高く評価いただいた施策は、基本的に継続して実施した。「データさきがけ」の研究者に対しては、第三者が有効に利用できる形でのデータの整備を進めるよう助言を行った。そして、(2)④で述べたように、データの収集・整備が進んだ。領域中間評価以降に、これらのデータを用いて、確認できている範囲で 15 編の論文が出版されている（領域中間評価時の実績は 1 編）。海外との連携では、領域中間評価時には計画段階であった、EU の研究者との交流を目的としたイベントを 2018 年にパリで開催し、領域中間評価以前から行っていた NSF との研究者の交流は継続して実施した。ただし、2020 年度にはコロナ禍が発生したため、従来と同様の交流は不可能となった。合同領域会議や公開シンポジウムをオンラインで開催したが、欧米とは時差が大きいため、合同領域会議の一部のセッションに国外から遠隔参加頂くにとどめざるをえなかった。日米欧からの同時参加は、どの時間帯に設定しても、どこかが深夜になるため断念した。

② 領域中間評価で今後の期待が述べられた事項への対応

領域中間評価において、本研究領域のマネジメントに関して今後の期待が述べられた事項としては、以下のものがあつた。

- ・ 「研究課題間あるいは複合領域内連携、および国際的連携へのさらなる取り組み」と「ビッグデータ法制度分野の研究成果のさらなる国際的な発信」（さきがけ）
- ・ 「ビッグデータ法制度分野、画像処理技術分野へのさらなる展開」と「5 コア分野の対象研究課題領域間での連携による“研究情報基盤の整備”」（CREST）

さきがけに関しては、領域中間評価後に研究期間が残っていたのは3期生のみであり、期間も2018年度の1年間のみであった。本研究領域は当時既に縮小段階に入っていたこと、ビッグデータ法制度に関する課題は終了済みであったことから、主として国際連携や国際的リーダーシップを重視すべきというご指摘と解釈して、「さらなる取り組み」に力を入れることとした。そして、国際連携のさらなる取り組みとしては、(2)③で述べたEUの研究者との交流イベントを新たに始めた。また、国際的なリーダーシップについては、2018年に募集が始まり、国際的プレゼンス向上の観点から質の高い国際連携を必須の要件としたAIP加速課題への展開を勧めることとした。これは、CRESTまたはさきがけの研究成果をベースに、新規のチーム型研究課題を提案するものである。2018年度の募集（研究開始は2019年度）は、CREST課題と同じ土俵で競う必要があったため、さきがけ研究者にとってはハードルが高かったが、3期生の島田と山田が採択された。なお、同年度に他に採択されたのは、ビッグデータ基盤とビッグデータ応用のCREST課題から各1件であり、両名は、AIPネットワークラボ内で非常に高く評価されたと考えられる。

CRESTにおけるビッグデータ法制度分野での活動は、主に山名チームが担当した。同チームに対しては、個人情報保護法等を単に分析・解釈するのではなく、制度設計者に対する提言ができるように努力することを勧めた。そして、同チームの研究者は、一般財団法人情報法制研究所等の研究者と連携し、暗号化されたデータの扱いに関する提言を行なった。2020年の個人情報保護法の改正で新設された「仮名加工情報」は、この提言の趣旨に沿ったものである。

CRESTの画像基盤技術分野に関しては、コンピュータビジョンに機械学習を適用する研究等が、CREST原田チームを中心に活発に行われた。CREST課題全体では、この分野のトップカンファレンスであるCVPRで15編、ICCVで6編、ECCVで2編の論文発表が行われた。領域中間評価後には、採択率2%程度のCVPRのオーラルセッションでの発表も行われている。国際的に見ても、この分野でトップクラスの研究成果を発信できたと考えている。

研究情報基盤の整備については、5つのコア分野の成果を完全に統合するには至っていないが、成果の一部を研究に活用できる情報基盤の整備は進んでいる。CREST課題の研究者に対しては、イノベーションに向けた取り組みに関する助言を随時行っており、4つのコア技術分野の成果は、論文等の文書だけではなく、ソフトウェア、データ、システム等の形でも公表・公開されている。各分野の研究者が利用できるもの、既に実際に利用しているものも多い。数理基礎技術分野とセキュリティ基盤技術分野ではオープンソースのライブラリ等として成果を公開している。これらの中には、CREST津田チームの並列多重検定ライブラリMP-LAMPのように数千コアでの大規模並列処理の実績があるもの、CREST山名チームの完全準同型暗号ライブラリ群のようにシステムの各レイヤでの最適化手法を統合したものなどが含まれる。自然言語・画像基盤技術分野でも、自然言語処理や画像処理のためのオープンソースのライブラリを公開しており、さらに、CREST黒橋チームでは、コーパスや学習済みモデルの公開も行っており、多数のダウンロード実績がある。また、CREST原田チームのクラ

中間学習の手法は、データ解析の現場で利用されるようになってきている。システムアーキテクチャ基盤技術分野でも、ビッグデータ処理向けのオープンソースのソフトウェア群を公開している。さらに、CREST 松岡チーム、合田チームのメンバーが所属する組織には、我が国の学術コミュニティ向けに大容量ネットワークと大規模計算資源を提供する国立情報学研究所および大学のセンター群が含まれることから、それら組織での実運用や実運用に向けた取り組みも進んでいる。なお、第5のコア分野であるビッグデータ法制度分野のCRESTにおける成果は、セキュリティ基盤技術分野の成果を普及させるにあたっての社会的課題を、法制度面から検討するものである。そのため、セキュリティ基盤技術分野と一体的に活動が行われている。このように、各コア分野の成果を学術研究において利用可能とする取り組みが進んだ。

(6) 人材育成

(4)①で述べた通り、CREST の総研究費に占める人件費の比率を高く設定することで、多くの若手研究者を雇用し、CREST 研究に参画する形で人材育成が行われた。さらに、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た。本研究領域から、2016年度は10名、2017年度は13名、2018年度は10名、2019年度は8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～2018年度にAIP ネットワークラボのCREST 領域（2016年度は3領域、2017年度は4領域、2018年度は5領域）から採択された若手研究者に対しては、成果報告会を兼ねて、一堂に会してポスター発表等で意見交換を行う場が設けられ、異分野連携に向けた訓練の場としても活用することができた（2019年度はコロナ禍のため中止となった）。ポスター発表が行われた2016年度～2018年度には、参加者の投票により、優秀な研究に対して毎年度ラボ長賞が与えられ、本研究領域が推薦した胡緯華（2016年度CREST 原田チーム）、包含（2017年度CREST 原田チーム）、榎田洋介（2018年度CREST 山西チーム）が各年度のラボ長賞第1位の表彰を受けた（2016年度と2017年度は単独1位、2018年度は同点で1位が2名）。

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きた。17名のさきがけ研究者の採択時と現在の身分の内訳は下の表のようになる。ほとんどが大学の教授または准教授に就任しており、それ以外を含め、順調にキャリアパスを歩んでいることが見て取れるであろう。

	採択時	2020年12月現在
教授	0	5
准教授	5	10
講師	2	0

助教	7	0
研究員	3	0
ユニットリーダー (国立研究開発法人)	0	1
執行役員 (ユニコーン企業)	0	1

(2)①でも述べた通り、1期生の松谷、3期生の島田、柳澤、山田がJSTのチーム型研究課題（CREST課題とAIP加速課題）の代表者として現在研究活動を行なっている。また、1期生の宮尾が2015年に日本学術振興会賞を、2期生の河原が2017年に文部科学大臣表彰を受賞した。研究期間終了後だが、3期生の山田も2020年に文部科学大臣表彰を受けており、それぞれに今後の活躍が大いに期待できる。

7. 研究領域として戦略目標の達成状況について研究を実施した結果と所見

(1) 研究総括のねらいに対する研究成果の達成度

「3. 研究総括のねらい」で述べた5つのコア分野（3(1)参照）のそれぞれで活発な研究活動が行われた。本研究領域のCREST研究全体では、発表論文が1,618編（領域中間評価時には911編。括弧内の数値は、以下同様。）で、そのうち国際論文は1,502編（850編）、口頭発表が2,491件（1,152件）で、そのうち招待講演は935件（570件）、特許出願が79件（47件）で、そのうち国際出願は47件（24件）の成果があった。さきがけ研究全体では、発表論文が118編（93編）で、そのうち国際論文は89編（68編）、口頭発表が443件（307件）で、そのうち招待講演が179件（118件）、特許出願は9件（5件）で、そのうち国際出願は1件（0件）の成果があった。なお、領域中間評価後の成果の増分比率で、さきがけがCRESTに比べて小さく見えるのは、領域中間評価後に残された研究期間が短いためである。

全発表論文のうち、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研究からは179編（72編）、さきがけ研究からは19編（11編）に達した。なお、これらの値は、「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ものである。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22編（15編）、さきがけ研究から1編（1編）であった。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。特筆すべき個別の研究成果については(3)(4)で述べる。

(2) 研究領域全体として見た時の特筆すべき研究成果

本研究領域の運営にあたって見据えていたのは、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、本研究領域の研究成果を俯瞰すると、次のような成果があがった。

- 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) 科学的・技術的な観点からの貢献

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、第5のコア分野である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うものであった。本節の「科学的・技術的な観点からの貢献」としては取りあげないが、暗号技術を適切に用いて、事業者にとってのマネジメントの手間・コストの削減につながる制度のあり方を提言することで、技術普及を支援するものであった。

(i) 数理基礎技術分野

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。

CREST 宇野チームは、粒度が比較的揃ったクラスタを網羅的に発見し、かつ再現性・一貫性も高いデータ研磨クラスタリング手法を開発し、収束性や計算量を理論的に評価した。若干の揺らぎで結果が大きく異なるようなクラスタリング手法では、実用上、信頼性に欠ける。また、得られたクラスタの粒度や抽象度の差が大きかったり、数が非常に多かたりすると、やはり扱いにくい。このような点に配慮した研究である。また、関連した列挙アルゴリズム等に関する研究も行なった。一連の研究成果は、トップカンファレンスのSTOC、KDDなどで発表されている。

CREST 加藤チームは、劣線形パラダイムを提唱し、構造の制約やサンプリングを活用したアルゴリズム、圧縮等の技術を活用したデータ構造、量子アニーリングを含む高速解法に問題をマッピングするモデリング技術などの分野で非常に多くの成果を生み出した。巨大デ

ータを扱う場合に、計算量のオーダーがかなり低くないと実用上計算不可能であることは自明であり、このような観点でアルゴリズムの専門家チームが課題に取り組んだことは極めて有益であった。チーム全体で400編以上の論文を発表しており、その中にはSODAなどのトップカンファレンスやアルゴリズム分野等の一流の論文誌で発表されたものが含まれる。代表的な成果を2つあげると、まず、ある種の階層的なクラスタ性を持つグラフは、グラフのサイズに依存せずに中心性などの性質を定数時間で検証可能であることを証明した。また、タンパク質に他の分子が結合することによって生じる自由度変化とその範囲を、高速に計算する数理的アルゴリズムを剛性理論に基づいて開発した。分子動力学法等のシミュレーション手法よりはるかに高速であり、酵素触媒の効能の謎に迫る成果である。この結果は、Scienceなどに掲載された。

CREST 原田チームは、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手法の開発とその理論的背景を明らかにした研究を嚆矢に、教師あり学習と遜色のない性能を理論的に保証できる弱教師付き学習に関する一連の研究を行なった。論文はNIPS/NeurIPS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。この中には採択率1%程度のNIPSのオーラルセッションでの発表も含まれている。この分野で世界の先駆けとなる成果であり、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きている。科学技術の進歩に大きく貢献していると言えるであろう。

CREST 津田チームは、高次元のデータから統計的に有意な仮説を探し出すために網羅的な探索と検証を試みる場合、優位性の基準を厳しくしないと偽陽性が多発し、厳しくすると偽陰性により発見が困難になる問題に取り組んだ。例えば、癌患者のさまざまなデータから癌の原因となる統計的に有意な複数要因の組み合わせを探するような場合に、こういう問題が発生する。同チームでは、選択的推論(Selective Inference)の一般的な枠組みのもと、仮説選択に対する制約を多数の線形不等式で表現し、仮説を選択的に生成して検証する方式を考案した。また、画像のセグメンテーションなどのアルゴリズムで計算結果と同時にp値(すなわち、どれくらい偶然ではないか)を求める方式も考案した。これらの成果は、ICML、CVPRなどのトップカンファレンスで発表されている。機械学習の手法を科学のさまざまな分野で活用する時代に、科学が科学であるためには、学習・推論結果の信頼性の客観指標が必要になる。この研究はそのような目的に資するものであり、科学を支える基盤として今後重要なものに発展することが期待できる。

(ii) 自然言語・画像基盤技術分野

CREST 黒橋チームは、本研究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、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た。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も実現した。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。

CREST 原田チームは、動画像を含む画像データの要約・理解・生成などに関する広範な研究を精力的に行った。代表的なものとしては、教師データが存在するソースドメインから教師データが存在しないターゲットドメインへの適合を、カテゴリ情報を反映して行う転移学習に関する研究がある。この成果は画像認識分野のトップカンファレンスである CVPR のオーラルセッション（採択率約 2%）で発表されている。人間が五感から得る情報の多くは視覚情報であり、言語とともに画像を用いた知的作業の可能性が広がることで、科学の発展に寄与することが期待できる。

(iii) セキュリティ基盤技術分野

CREST 佐久間チームは、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、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 1,000 倍近くの高速度化、メモリ量で 50 万倍以上の効率化を達成したものであった。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。

CREST 山名チームでも、完全準同型暗号の既存のライブラリである HElib より桁違いに高速なライブラリの開発とそのための要素技術に関する研究を行った。理論とシステムアーキテクチャの両面から検討を行い、SIMD 化、メモリ割り当ての最適化、表検索の導入などによる高速化が図られた。既存のものと比較した高速化の倍率は、アプリケーションによって異なるが、データマイニングなどの応用で数百倍の高速化を実現した。

(iv) システムアーキテクチャ基盤技術分野

CREST 松岡チームは、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 月から連続で世界一の成果をあげ続けた。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、広範な科学技術の分野で利用可能な強力なツールを提供するものである。代表者の松岡は、現在、我が国のフラッグシップスーパーコンピュータ富岳を運用する理化学研究所

計算科学研究センターのセンター長を務めており、科学技術の進歩に大きく貢献することが期待できる。

CREST 合田チームは、クラウド上に科学研究のための計算実験環境を迅速かつ再現性をもって構築できる方式に関する研究を行なった。人手での設定の手間を削減し、構築を容易に行えるようにするために、資源割当の最適化を自動化する方式の研究も行なった。これらの基本的な方式については IEEE Cloud、IEEE Big Data などでも発表されている。また、ゲノム解析に特化する形で、この分野の標準的なワークフローの枠組みを用い、資源割当の最適化のための情報を自動取得する方式の研究も行なった。今日、さまざまな科学分野でデータ解析やシミュレーションが活用されるようになり、その比重も次第に重くなっている。特定のクラウド事業者依存せず、資源割当の自動化等による環境設定の省力化、迅速性、再現性を実現することには意義がある。このチームは、大学共同利用機関と共同利用・共同研究拠点に所属する研究者が主要なメンバーであり、広範な科学研究を実際に支援することが期待できる。

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た。また、本研究領域の特徴的活動である「データさきがけ」の研究も行なった。代表的な成果を以下で簡単に説明する。

(i) 数理基礎技術分野

1 期生の佐藤は、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。そして、この結果に至る一連の成果を、トップカンファレンスの KDD、ICML で発表した。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。

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った。そして、成果を、トップカンファレンスの KDD や ISMB などでも発表した。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た。さらに、研究期間終了後の 2018 年度～2020 年度には AIP 加速 PRISM 研究「創薬標的分子の確からしさを検証するツール物質の探索」に研究分担者として参加し、圧縮技術を用いた研究を継続して行なった。

佐藤と田部井は、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ることで、科学・技術の発展に貢献した。一方、2 期生の杉山は、基礎的な手法で優れた成果をあげた。まず、データから統計的に有意に出現するパターンを発見する統計的パターンマイニングの基礎的な技術を確立し、さらに、情報幾何の概念を用いて、パターン集合の空間の確率モデルの学習を部分多様体への射影とみなすことで、既存手法より桁違いに高速なアル

ゴリズムを考案した。これらの研究の成果は、KDD、NIPS、ICML などのトップカンファレンスで発表した。

(ii) 自然言語・画像基盤技術分野

自然言語基盤技術に関しては、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で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。一連の成果を、自然言語分野のトップカンファレンスである ACL、COLING、EMNLP などで発表した。自然言語、画像、データベースという人間が知識を表現する主要なメディアのビッグデータを、統合的に活用するための基盤技術に関する研究であり、科学の発展に大きく貢献する可能性を秘めている。

画像基盤技術に関しては、2 期生の大竹、3 期生の酒向らが応用分野の科学・技術の発展に貢献する成果をあげた。大竹は、CT 等の医用画像情報から、骨盤と大腿骨の位置推定と動体解析を行う研究、筋骨格モデルを構築する研究などを行なった。これらは診断、手術計画、術後の長期的な予後予測などに利用可能な成果である。筋骨格モデルの構築に関する研究では、医用画像分野のトップカンファレンスである MICCAI (International Conference on Medical Image Computing & Computer Assisted Intervention) で Best Paper Award を受賞した。

酒向は、宇宙観測に動画像を用いるという先駆的なテーマに取り組み、微光物体や高速移動物体を動画像解析技術で検出する方式などの研究を行った。そして、いくつかの短時間で変動する現象の観測に世界で初めて成功した。これらの現象には、連星ブラックホール、高速電波バーストなどに関して天文学に新たな知見をもたらすものが含まれる。さらに、スペースデブリの地上からの観測など実用的な価値がある成果もあげた。科学的・技術的に、大きな可能性を新しく切り拓いた研究である。

(iii) システムアーキテクチャ基盤技術分野

1 期生の松谷は、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る。今日、Intel や AMD が、CPU、GPU、FPGA を統合的に扱える企業を目指して動き始めているが、その先にある未来のビッグデータ処理基盤を見据えた研究と考えることもできる。

(iv) ビッグデータ法制度分野

1 期生の生貝は、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い、新聞への寄稿などの一般向けの情報発信も行った。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。

(v) データさきがけ

3 期生の小野木、柳澤、山田がデータの収集と整理を進めた。小野木は、全国規模で 1961 年～2015 年の 55 年間にわたる大豆の栽培データと気象やゲノムなどのそれに関連するデータの収集・整理を行った。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を 400 人日分以上収集・整理した。山田は人の腸内環境データとそれに対応した内視鏡所見や生活データを約 2,800 検体収集・整理した。

さきがけの研究期間終了後も収集・整理は続いている。小野木は 2019 年までのデータを新たに加え、さらに農林水産省の研究推進事業において他のデータとの統合やユーザインタフェースの充実にも取り組んでいる。柳澤が収集・整理したデータは 700 人日分以上まで増加しており、山田が収集・整理したデータも約 4,000 検体まで増加している。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。データを分析することで得られた結果の一例をあげると、山田のデータからは、腸内細菌叢と大腸癌のステージとの相関などが得られている。これは、大腸癌の早期発見などに役立つ可能性がある成果である。

(4) 社会的・経済的な観点からの貢献

(i) CREST 数理基礎技術分野

CREST 山西チームは、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た。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を進めた。これは、OCT(Optical Coherence Tomography)を用いて網膜の厚みを測り、その結果から進行予測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教室の協力を得て行われた。

CREST 加藤チームでは、大規模な道路ネットワークでの避難計画策定に、数理最適化を適

用する研究を行なった。これは、南海トラフ地震が発生した場合の大阪市を対象に、津波での浸水を考慮した避難計画問題を解くものである。現実の大都市の複雑な状況を想定したものであり、粗視化を行うことで、数理最適化手法を適用可能とし、現実的な時間で問題を解けるようにした。自然災害も少子高齢化と同様に、我が国の社会に突きつけられたもっとも大きな問題群の一つである。このような問題に対する取り組みが着実に行われた。

(ii) CREST 自然言語・画像基盤技術分野

CREST 黒橋チームは、企業や自治体と協力した取り組みを行った。一例をあげると、(株)インサイトテック社が集めた700万件超(研究時)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、本研究領域の研究成果が活用された。また、別の事例として、LINE(株)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も行われた。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。また、CREST 黒橋チームの研究成果を基に、AIP 加速 PRISM 研究「疾患知識ベースの構築と医療テキストの知識処理」が実施されており、医療分野のテキストを対象とした研究にも発展している。

CREST 原田チームは、2つの教師データをランダムな比率で混合して新たな教師データを生成することで、教師データが少ない場合でも、高い分類性能を持ったモデルを学習できるクラス間学習手法を考案した。そして、画像認識や環境音認識での有効性を示した。既にKaggleのコンペティションなどでも一般的に使われるようになっており、実用性が高く使いやすいことから、社会的・経済的な貢献が期待できる。

課題研究期間終了後のことになるが、2020年のコロナ禍において、世界各地の公式サイトやニュースを収集し、国・地域とカテゴリで集約して日本語と英語で提示する「COVID-19 世界情報ウォッチャー」(<https://lotus.kuee.kyoto-u.ac.jp/NLPforCOVID-19/>)が迅速に立ち上げられた。ここでは、自然言語基盤技術分野の研究活動を行なったCREST 黒橋チームの黒橋と乾、さきがけ1期生の宮尾、同2期生の河原が重要な役割を果たした。

(iii) CREST セキュリティ基盤技術分野

CREST 佐久間チームは、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、病院や遺伝子検査施設と協力して、生活習慣病の予防医療を支援する実証実験を行なった。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。

CREST 宮地チームは、独自の秘密分散方式により、各学校に分散された事故に関する情報

を、他の学校等に漏れないように機密保持した状態で統合的に分析する実験を行なった。複数の医療施設や学校のデータを統合することで、単独の医療施設や学校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ような実験には、大きな意義がある。

(iv) CREST システムアーキテクチャ基盤技術分野

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た。また、CREST 合田チームが開発した技術の一部は、国立情報学研究所のサービスで利用され、TSUBAME3.0 などでも実運用されている。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。

(v) さきがけ

3 期生の島田は、九州大学で実際に行われている授業の学習データを分析・可視化することで、授業前の予習状況や授業中の学習状況の把握、教材の要約の自動生成とそれを用いた予習・復習支援、教材改善の支援などに関する研究を行った。これは、実データを用いた実証実験であり、教育・学習という社会の根幹を支える活動を対象にしているため、社会に大きく貢献できる潜在力を持っている。

3 期生の山田は、大腸癌患者の便に対するメタゲノム・メタボローム解析を行い、癌の進行段階に応じ増減する細菌や代謝物質の同定を行った。これは、大腸癌の早期発見や予防につながる可能性がある研究成果である。山田はさきがけの研究テーマであった腸内細菌に関連して、「日本人腸内環境の全容解明とその産業応用プロジェクト」(JCHM: Japanese Consortium for Human Microbiome) を立ち上げて代表を務めるとともに、研究開発型ベンチャー企業の株式会社メタジェンで取締役副社長 CTO としても活躍している。この分野で社会的・経済的に大きく貢献する可能性が高い。

(5) 本研究領域に続く研究資金の獲得状況

AIP ネットワークラボに所属する領域の研究課題のうち、2018 年度以降に終了するものについては、研究期間中に得られた優れた成果をもとに新たな方向づけを行うことで、AIP 加速課題として採択される道が拓かれた。本研究領域で対象となりうる CREST 10 課題、さきがけ 6 課題のうち、CREST から 3 課題、さきがけから 2 課題が採択され、研究資金を獲得して現在研究を遂行中である。これに加えて、AIP 加速 PRISM 研究に CREST から 2 課題が採択された。これらはすべてチーム型の研究課題であり、CREST やさきがけに既に採択された課題が対象となるため、競争はきわめて厳しいものである。

さらに、さきがけ研究者からは、他の CREST 領域で代表者として課題に採択されたものが

2名、主たる共同研究者として採択されたものが1名いる。

8. 総合所見

(1) 研究領域のマネジメント

まず、5つのコア分野を設定し、コア分野間のバランスを考えて課題を採択した。課題の選考にあたっては、高い独創性・新規性、コラボレーションの可能性、提案者の明確な目的意識などを重視した。各課題の実施期間中は、領域会議、評価会、サイトビジットなどを組み合わせ、課題の進捗の把握と助言に努めた。そして、領域内外との連携、特に国際連携を促進する施策を重視し、「データさきがけ」のような新しいタイプの課題の実施にも挑戦した。

(2) 戦略目標の達成状況

戦略目標の2つの達成目標のうち、本研究領域で達成を目指したのは「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」の方である。この「次世代基盤技術」に関して、以下のように大別される多くの研究成果を生み出すことができた。

- 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

前半の2項目は、統合解析を行うプラットフォームが提供するサービスの高速化と高度化を担うものであり、後半の2項目は、プラットフォームの高速化・大容量化と強靱化を担うものである。個別技術の「創出・高度化」に加え、知識インフラの体系化（CREST 黒橋チーム）、計算量による体系化（CREST 加藤チーム）、潜在変数モデルの変化検知・変化予兆検知のための系統的な方法論（CREST 山西チーム）、多数の属性を持つ高次元データの統計処理に関する系統的な方法論（CREST 津田チーム）などの成果も得られている。

一方、「次世代基盤技術」の利用目的である「様々な分野のビッグデータの統合解析」の観点では、まず、「データさきがけ」で収集・整理されたデータは、3件とも複数の分野のデータを紐づけたものであり、「様々な分野のビッグデータの統合解析」を当然の前提としたものである。これら以外にも実データを用いた様々な解析の実験が行われた。そして、統合解析では、異なる分野の多数の属性を紐づけた解析が行われ、高次元データの統計処理を要することが多く、そのような処理結果の信頼性を評価する方式についても CREST 津田チームが重要な成果を得ている。

以上をまとめると、本研究領域が達成すべき目標である「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」を目指し、様々な角度からの

挑戦を行い、最終的に、目標を達成する十分な成果を得ることができた。

(3) 本研究領域を設定したことの意義

データが価値の源泉であることは、産業分野でも学術分野でも認識されるようになってきている。ビッグデータ基盤技術が、今後の社会問題の解決や学術研究の発展に欠かせないことは明らかである。ビッグデータ基盤技術をテーマとした研究領域が設定されたこと自体は、必然であったと言っても良いであろう。

本研究領域が設定されたことはもちろん、それが 2013 年度であったことの意義も大きい。米国においてビッグデータイニシアチブに関する公告が発表されたのが 2012 年であり、諸外国に遅れることなくビッグデータの研究プログラムを開始できたことになる。当時は、まだ第 4 期科学技術基本計画の時代であり、第 5 期科学技術基本計画の中に Society 5.0 のようなビジョンが現れたのは、本研究領域が設定された 3 年後のことであった。本研究領域で力を入れた米国 NSF との連携が円滑に行えたのも、日本に早い段階からカウンターパートが存在すると、先方が認識してくれていたからである。そのおかげで、この分野での組織的な連携を、遅れをとることなく始めることができた。また、研究期間が 2013 年度～2020 年度に設定されたことで、ちょうどこの時期に起こった急激な技術変化による強い刺激を受けながら、研究活動を進めることができた。ビッグデータの収集・蓄積が進み、GPU やクラウドの発達による利用可能な計算能力も急増し、それを追い風とした深層学習などの機械学習技術が急激に発展した時期に、本研究領域の活動が行われたことは、非常に有意義であった。

(4) 科学技術イノベーション創出に向けた、今後への期待、展望、課題

さまざまな分野において、科学のデータ駆動化が進んでいる。生命科学では大量のオミクスデータの取得が可能となり、地理情報分野ではドローンなどによる高精細なデータ取得が可能となった。本研究領域でも、天文観測データを静止画から動画に変えることで、データ駆動科学をさらに高度化する研究なども行われた。本研究領域の活動開始後にも、科学のデータ駆動化を進める活動は世界中で行われている。たとえば、データ自体を出版対象としたジャーナルとして有名な Nature Scientific Data が創刊されたのは 2014 年であり、本研究領域の発足後のことである。

以上のような時代背景のもと、科学技術イノベーションに不可欠な科学の発展自体が、ビッグデータ解析技術に大きく依存するようになった。本研究領域でも、医学、生命科学、化学、天文学、農学、経済学、教育学などの分野のデータを収集・解析する研究が行われ、これらの一部ではあるが、科学としての新しい発見にもつながっている。これらの中には、純粋科学の発展だけでなく、社会的な波及効果が期待できるものも少なくない。これらに加え、分野を限定しない解析手法に関する研究も進められた。その中には、ビッグデータ解析の再現性や検定のような、ビッグデータ基盤技術が科学の基盤となるために避けて通れないテ

ーマに関するものもある。また、汎用性の高い高速解析技術のように、今後どのように社会に浸透するか予測しづらいが、イノベーションへの展開が期待できるものもある。

この節の最後に、研究成果の継続的な利用に関する課題を述べる。科学研究の成果が主に論文であった時代には、その成果は図書館にアーカイブされて活用された。図書館の運営費は研究費とは別に手当てされており、成果利用の継続性が図られていた。しかし、人間が読む情報だけでなく、機械可読なデータやソフトウェアが科学の発展にとって重要な成果となってきた今日では、これらのメンテナンスを誰がどうやって行うかが問題となる。研究の過程で取得・作成されたデータやソフトウェアは、論文を書いたら用済みとなるようなものであってはいけない。これらは原理的なコピーのコストがほぼゼロであり、第三者が利用可能な形で提供し、第三者と共有することで、価値が増大するものである。科学技術イノベーションのためのデータを中心としたエコシステムを構築する必要がある。これは本研究領域だけで解決できる問題ではなく、今後の科学政策で考慮すべき問題であろう。

(5) その他の所感

2020年はコロナ禍に見舞われ、オンライン化やデジタル化が急激に進んだ。企業の在宅勤務や大学のオンライン授業などで、組織によっては、「数週間で10年分は進んだ」と感じているところもあるだろう。我々を取り巻く大きな社会問題の中には、少子高齢化や気候変動などのように10年単位での予想が比較的行いやすいものもあれば、突発的な自然災害やパンデミックのように一瞬先のこともよくわからないものもある。一方、本研究領域のCREST課題は5.5年、さきがけ課題は3.5年の研究計画を提案するものであった。研究のデジタル化が研究サイクルの短縮に貢献し、変化のスピードが早くなった現代において、数年単位の計画を大きく外すことなく立てることは、挑戦的な研究であればあるほど難しい。研究計画の変更が頻繁に発生する方が、自然となりつつある。しかし、研究計画の中には簡単に変えてはいけないものも含まれる。例えば、研究者の雇用計画が頻繁に変更されたら、落ち着いて研究を行うことができなくなり、ネガティブな影響の方が大きくなるだろう。したがって、長期的なウォーターフォール型の計画と短期的なアジャイル型の計画をうまく組み合わせることが必要になると思われる。

一方、(4)でも述べたように、図書館のアーカイブ機能のように長期的な利用を保証する仕組みが、研究成果としてのデータやソフトウェアにも求められるようになっている。これは、課題の研究期間より長いスパンで考えた時に発生する問題である。急激な変化への対応と長期的な安定性という両極端の問題に対して、個々の研究課題ではなく、研究コミュニティ全体や社会全体にとっての最適解を見つける努力が必要と思われる。

以上

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」

複合領域事後評価用資料

研究総括：喜連川 優

副研究総括：柴山 悦哉

2021年12月

(新型コロナウイルス感染症の影響を受けて研究期間の延長された
CRESTの2研究課題に対する追記)

(2021年12月 追記)

新型コロナウイルス感染症の影響を受けた以下の2課題について、2021年9月30日まで、半年間の研究期間の延長を行った。

・「ビッグデータ時代に向けた革新的アルゴリズム基盤」

研究代表者:加藤 直樹 (兵庫県立大学 大学院情報科学研究科 研究科長・学部長)

・「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」

研究代表者:山名 早人 (早稲田大学 理工学術院 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」については、2020年度に開始した新型コロナウイルス感染症の予後予測および同ウイルスの分子レベルでの感染メカニズムに関する研究等を継続して行なった。

その結果、入院患者が重症化するかどうかをCT画像と臨床情報から予測する機械学習モデルで、2020年度の成果と比較して、同程度の精度でパラメータ数がより少ないものを構築することができた。また、新型コロナウイルスのスパイクたんぱく質が人間の受容体に結合して感染を引き起こす際の動的振る舞いを、本課題で研究に取り組んできた剛性理論を用いたアルゴリズム等を用いて明らかにすることができた。

延長により、本課題の革新的アルゴリズム基盤が、新型コロナウイルス感染症のような突発的大問題に迅速に対応するための重要な技術資産となりうることを、より強く確信できるようになった。これは、今後のイノベーションへの貢献がより一層期待できる成果である。

「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」については、当初の研究期間中に実施が困難となった実証実験を行なった。また、公開ライブラリの拡充等も行なった。

その結果、リストバンド型センサデバイスを用いて、実験参加者から1万人日分を超えるライフログデータの収集を行い、データを暗号化したまま生活比較分析を行うことができた。また、実験参加者の許諾を得た4,200人日分のデータを公開することもできた。さらに、薬局をフィールドとした実証実験では、顧客のパーソナルデータを秘匿した医薬品副作用解析をテーマに4店舗での実験を行い、サービス提供の可能性を示すことができた。

延長により、現実の問題への適用可能性の検討・検証が進み、今後のイノベーションに向けた展開をより一層後押しする成果が得られた。

今回、「ビッグデータ時代に向けた革新的アルゴリズム基盤」の研究者が本研究領域で積み重ねた研究成果は、新型コロナウイルス感染症という新たな問題への迅速な対応を可能にするものであった。これ以外にも、本研究領域の終了した研究課題の研究者が、「COVID-19世界情報ウォッチャー」などのプロジェクトで、新型コロナウイルス感染症への対応をおこなっている。これらの研究活動や成果の社会への還元実績は、本研究領域の意義を改めて示

すものである。新型コロナウイルス感染症に限らず、様々な社会問題の解決のために、本研究領域での研究成果や経験が活かされることを期待する。