

## 科学研究費助成事業 研究成果報告書

平成 30 年 6 月 6 日現在

機関番号：12601

研究種目：新学術領域研究（研究領域提案型）

研究期間：2013～2017

課題番号：25120001

研究課題名（和文）スパースモデリングと高次元データ駆動科学創成への支援と広報

研究課題名（英文）Support and publicity for initiative for High-dimensional Data-Driven Science through deepening of sparse modeling

研究代表者

岡田 真人 (Okada, Masato)

東京大学・大学院新領域創成科学研究科・教授

研究者番号：90233345

交付決定額（研究期間全体）：（直接経費） 59,000,000円

研究成果の概要（和文）：領域代表が主導となり、多様な分野の連携を効果的に促すために、データ駆動科学の三つのレベルを提唱した。データ獲得を行う「計算理論」のレベルと、アルゴリズムや性能評価を担当する「表現・アルゴリズム」のレベルの間に、モデリングのレベルが必須であることを見出し、モデリンググループ(B01)が中心となり、計画研究間や公募研究との強固な連携体制が整った。この体制のもと、自然科学者と数理情報科学者が緊密に連携することで、分野や階層を超えた多数の共同研究が生まれ、多数の学術論文として結実している。また、積極的なアウトリーチの結果、幅広い分野・社会・産業界に、スパースモデリング・データ駆動科学が浸透しつつある。

研究成果の概要（英文）：In order to effectively promote collaboration in various fields, regional representatives are leading the three levels of data driven science. We found that the level of modeling is essential between the level of "computational theory" that acquires data and the level of "expression / algorithm" in charge of algorithms and performance evaluation. The modeling group (B 01) is the main part of this project, and strong collaboration system between planned research and public research has been established. Under this system, by collaborating closely with natural scientists and mathematical information scientists, numerous collaborative research beyond fields and hierarchies has been born and fruited as a large number of academic papers. In addition, as a result of aggressive outreach, sparse modeling / data driven science is spreading to a wide range of sectors, industries, and industries.

研究分野：データ駆動科学

キーワード：高次元データ駆動科学 スパースモデリング データ駆動科学の実践 モデリング原理の確立 数理基盤形成 三つのレベル

### 1. 研究開始当初の背景

より深く自然を知りたいという飽くなき探究心が、とどまることを知らない計測技術の向上をもたらし、大量の高次元観測データを日々生み続けている。

これを好機と捉え、科学技術の水準を革新的に向上・強化させるために、情報科学と自然科学が緊密に融合した革新的な自然探究の方法論を構築する必要がある。我々は、その方法論構築のためのキーテクノロジーが、近年、情報科学分野で大きな注目を集めているスパースモデリング(SpM)であると考えた。SpMは、高次元データに普遍的に内在するスパース(疎)性を利用することで、計算量が次元数に対して指数爆発(計算量爆発)し、スパコン「京」でさえ数十年を要する状況でも、実際の時間でデータから最大限の情報を効率よく抽出できる技術の総称である。これまでも SpM は個別分野において萌芽的成果を生み出しており、それらの背後にある共通原理を明確化し、より強力なものにすることは自然科学全体に革新的展開をもたらすものと考えられる。

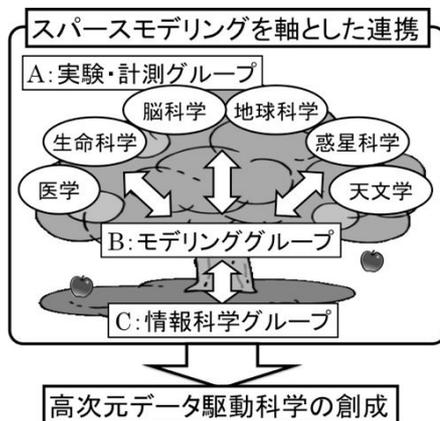


図 1: スパースモデリングに基づく高次元データ駆動科学

### 2. 研究の目的

こうした背景を踏まえ、本領域では図1のように、SpMや高次元データ解析で顕著な実績をあげている情報科学者と、生命分子からブラックホールに至る、幅広い自然科学の実験・計測研究者が SpM というキーテクノロジーを軸として緊密に連携することで、大量の高次元データを効率的に科学的な知へとつなげる高次元データ駆動科学ともいべき新学術領域を創成する。これにより、これまで、個々の分野ごとに探求されていた課題に対して、新たな共通原理に基づく革新的な科学的方法論を確立することによって、あらゆる科学分野の研究に大きな波及効果をもたらし、来たるべきデータ科学の時代に向けて、我が国の学術水準の圧倒的優位性を確立する。

### 3. 研究の方法

本領域は情報科学者と幅広い自然科学の実験・計測研究者が参加する。総括班はこの特長を活かし、領域メンバーの緊密な連携を促進させ、個々の成果がモデリング原理の確立を通じて高次元データ駆動科学の組成につながるように、領域全体を先導する。また、データ解析技術の導入による実験・計測分野の進展だけにとどまらず、高次元データ駆動科学の情報数理基盤の形成へつなげる。このような支援活動および広報活動に関連し、以下の三つの課題に取り組む。

#### 【課題1】データ駆動科学の実践

実験グループ(A01, A02)とモデリンググループ(B01)から数名の担当を配置し、各計画研究間の研究進捗の調整を行い、本応募領域の要である情報科学と自然科学の有機的連携を促進する。自然科学者と情報科学者が密に共同し、仮説の提案・検証ループを高速に回す研究の具体的規範例を各分野で実践する。

#### 【課題2】データ駆動科学の情報数理の基盤形成

情報科学グループ(C01)の持つ最先端の数理解情報科学を SpM に融合することで、種々の自然科学課題の状況に柔軟に対処できる新たな数理解析手法の構築につなげる。総括班としては、【課題1】で浮き彫りになった問題点を整理し、モデリンググループと情報科学グループを主導して問題解決にあたる。

#### 【課題3】データ駆動科学創成に向けた若手人材育成と広報・アウトリーチ活動

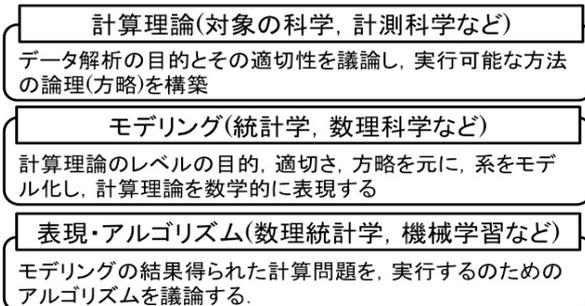
若手主体の研究集会や、若手による海外研究機関での共同研究やセミナー発表を支援する。異分野研究室間の連携を支援し、自然科学と情報科学の両方の素養を備えた人材を育成する。様々な学術広報活動を通じて、高次元データ駆動科学を全学術領域へ浸透させる。国際研究集会の主催もを行い、データ駆動科学のさきがけであることを世界に発信する。

### 4. 研究成果

岡田領域代表が主導となり、多様な分野の連携を効果的に促すために、図2に示すデータ駆動科学の三つのレベルを提唱した。データ獲得を行う「計算理論」のレベルと、アルゴリズムや性能評価を担当する「表現・アルゴリズム」のレベルの間に、モデリングのレベルが必須であることを見出し、モデリンググループ(B01)が中心となり、計画研究間や公募研究との強固な連携体制が整った。

この体制のもと、自然科学者と数理解情報科学者が緊密に連携することで、分野や階層を超えた多数の共同研究が生まれ、多数の学術

論文としても結実している。また、積極的な広報・アウトリーチの結果、幅広い分野・社会・産業界に、スパースモデリングおよびデータ駆動科学が浸透しつつある。以下に、膨大な研究成果から一部を抜粋して紹介する。



### 【課題 1】データ駆動科学の実践

データ駆動科学の実践を目的とする実験グループ(A01, A02)は, 生物学・地学などの理科第 2 分野を中心に構成されているグループである。モデリンググループ(B01)・情報科学グループ(C01)が開発した新手法群を利用することで, 個別分野では得られなかった新知見が得られつつあり, データ駆動科学の実践を進めている。

たとえば, スパースモデリングの基礎ツールである L1 正則化法について, 各計測への導入を計測モデリング班(B01-1)との連携を中心に行って顕著な成果を上げている。たとえば, 医学班(A01-1)では, RI 高速撮像への導入を進めているが, 既に放射線医が実際の現場で利用可能なレベルに至っている。また, 天文学班(A02-3)においては, ブラックホールの直接撮像に向けて, 様々な人工データを用いた検証に成功している。

スパースモデリング班(B01-2)が開発した全状態探索法は, L1 正則化法などと異なり, 考えられる全ての候補解を与えてくれることから, 潜在構造抽出を目的とする自然科学の課題解決にとって重要な手法である。市川公募班は, 同手法を NIRS 計測データに適用することで, 乳幼児の ASD(自閉症スペクトラム)児と ADHD(注意欠陥・多動性障害)児の判別を可能にした。また, 地球科学班(A02-1)では, 2011 年東北沖津波堆積物の判別および起源の推定を行っている。

また, スパースモデリング班(B01-2)が開発したベイズスペクトル分解法も, スペクトル計測の普遍性を背景に, 多様な分野に応用されている。たとえば, 生命科学班(A01-2)は, 二次元 NMR スペクトル画像に適用したほか, 惑星科学班(A02-2)では, 惑星の反射スペクトル計測データへの応用に成功している。

以上に挙げたスパースモデリング技術の他にも, 多数のデータ駆動型手法により各個別分野でデータ駆動科学の実践が行われて

いる。たとえば, 物理モデリング班(B01-3)と地球科学班(A02-1)の連携により, データ同化などの先進的手法が岩石形成のダイナミクス解析に導入されている。

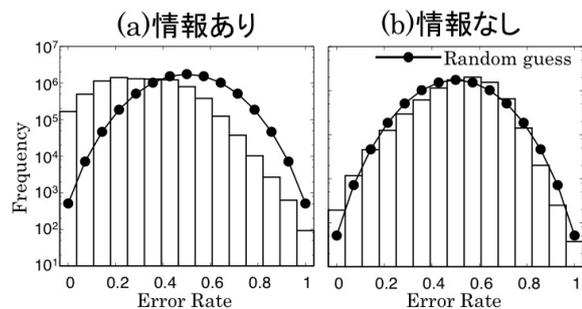
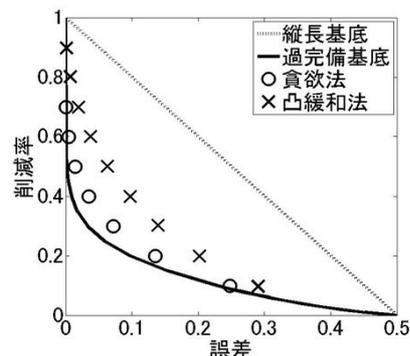


図 3: 与えられたデータに対する 2 クラス識別の情報有無の判別結果

### 【課題 2】データ駆動科学の情報数理の基盤形成

既存の SpM である LASSO では, データの潜在構造である説明因子の組み合わせが安定して求まらない問題があることを見出し, 次元に対して指数オーダーある説明因子の全組み合わせに対して求める SVM 全状態探索法を提案し, 図 3 に示すような状態密度関数(Density of State, DoS)を求めることがデータ駆動科学にとって本質であることを見出し, ES-DoS と名付けた。さらに, 指数オーダーの計算量がかかる問題についても, 交換モンテカルロ法を利用して効率的に状態密度関数を計算する近似的全状態探索法(AES-DoS)をスパースモデリング班(B01-2)が提案した。

また, 大自由度班(C01-3)とスパースモデリング班(B01-2)との協働により, 過完備基底に基づいた次元削減の性能について精緻な分析を行い, 過完備基底を用いるとランダム射影による素朴な方法よりも大幅に優れた性能を達成できることを示した(図 4)。また, 次元削減を行う実際的手法として貪欲法と凸緩和法を比較し, 貪欲法が優れていることを示した。この発見は, より有用なアルゴリズムの開発の指針を示したという点で, SpM の深化を目指す本領域にとって重大な知見である。



【課題 3】データ駆動科学創成に向けた若手人材育成と広報・アウトリーチ活動

本領域に参画した研究者は 985 人であり、そのうち 39 歳以下の若手研究者は半数近い 437 人である。本領域には数多くの若手研究者が参画し、次世代のデータ駆動科学を担う人材を着実に輩出した。その効果は、受賞やキャリアアップの件数、若手主体の学術論文数の多さなどに如実に表れている。

広報・アウトリーチ活動も活発に行った。2015 年 12 月および 2017 年 9 月に京都メルパルクにおいて開催した本新学術主催の国際シンポジウム「International Meeting on “High-Dimensional Data-Driven Science” (HD3-2015,2017)」では、国内外のデータ駆動科学のキーパーソンによる招待講演が行われたほか、若手研究者も多数講演を行った。これらの研究内容は、IOP Publishing のオープンアクセス誌 Journal of Physics: Conference Series において特集号として組まれている。さらに、それぞれの研究分野においても、データ駆動科学関連の研究集会・企画シンポジウムを多数開催し、データ駆動科学の普及・浸透がなされている。

社会・国民・産業界においても、広く普及活動を行っている。プレスリリース・テレビ・新聞・雑誌などによる広報のほか、一般も参加できるチュートリアルなど、多方面に対し、本領域の研究活動の広報を行った(図 5)。その結果、社会・産業界においてもデータ駆動科学とスパースモデリングは現場ですぐに使える方法として認識されている。実際に、公開シンポジウムやチュートリアル参加者の企業関係者割合は非常に高く(例えば 2018/3/31 主催の最終シンポジウムにおいて 50%以上)、アカデミックのみならず産業界への貢献の非常に大きい期待がうかがい知れる。



図 5 : NHK サイエンスゼロ(2015/5/3) . 本領域全体の研究内容が特集された .

5 . 主な発表論文等

(研究代表者、研究分担者及び連携研究者に

は下線)

(雑誌論文)(計 10 件)

1. \*R. Nakata, H. Hino, T. Kuwatani, S. Yoshioka, M. Okada and T. Hori, "Discontinuous boundaries of slow slip events beneath the Bungo Channel, southwest Japan", Scientific Reports, 7, 06185-1-7, 2017. (査読有) [海洋機構からプレスリリース]
2. \*P. K. Hong, H. Miyamoto, T. Niihara, S. Sugita, K. Nagata, J. M. Dohm and M. Okada, "An automatic deconvolution method for modified Gaussian model using the exchange Monte Carlo method: application to reflectance spectra of synthetic clinopyroxene.", Journal of Geology & Geophysics, 5(3), 1000243, 2016. (査読有)
3. 五十嵐康彦, 竹中光, 永田賢二, \*岡田真人, "AI for Science とデータ駆動科学 — ベイズ計測と VMA の提案 —", 応用統計学, 45(3), 75-86, 2016 . (査読有)
4. \*Y. Nakanishi-Ohno, M. Haze, Y. Yoshida, K. Hukushima, Y. Hasegawa and M. Okada, "Compressed sensing in scanning tunneling microscopy/spectroscopy for observation of quasi-particle interference", Journal of the Physical Society of Japan, 85(9), 093702-1-5, 2016. (査読有) [東大からプレスリリース]
5. Y. Igarashi, K. Nagata, T. Kuwatani, T. Omori, Y. Nakanishi-Ohno and \*M. Okada, "Three levels of data-driven science", Journal of Physics: Conference Series, 699, 012001, 2016. (査読有)
6. \*T. Kasai, K. Nagata, M. Okada and T. Kigawa "NMR spectral analysis using prior knowledge", Journal of Physics: Conference Series, 699, 012003, 2016. (査読有)
7. K. Nagata, J. Kitazono, S. Nakajima, S. Eifuku, R. Tamura and \*M. Okada, "An exhaustive search and stability of sparse estimation for feature selection problem.", IPSJ Transactions on Mathematical Modeling and Its Applications, 8(2), 23-30, 2015. (査読有)
8. \*T. Kuwatani, K. Nagata, M. Okada, T. Watanabe, Y. Ogawa, T. Komai, N. Tsuchiya, "Machine-learning techniques for geochemical discrimination of 2011 Tohoku tsunami deposits", Scientific Reports, 4, 7077-1-6, 2014. (査読有) [東北大学からプレスリリース]
9. \*H. Ichikawa, J. Kitazono, K. Nagata, A. Manda, K. Shimamura, R. Sakuta, M. Okada, M. K Yamaguchi, S. Kanazawa, R. Kakigi, "Novel method to classify hemodynamic

response obtained using multi-channel fNIRS measurements into two groups: Exploring the combinations of channels", Frontiers in Human Neuroscience, 8, 480-1-9, 2014. (査読有) [中央大学からプレスリリース]

10. \*M. Honma, K. Akiyama, M. Uemura, S. Ikeda, "Super-resolution imaging with radio interferometry using sparse modeling", Publications of the Astronomical Society of Japan, 66, 9514-1-12, 2014. (査読有)

[学会発表](計 51 件)

1. 岡田真人, "AI for science: データと計算量の爆発", 第125回東京大学公開講座(招待講演) 2017年.
2. 岡田真人, "AIで材料開発?", NIMS 一般公開 2017(招待講演) 2017年.
3. 岡田真人, "構造材料科学の三階層モデルとスパースモデリング", 日本金属学会 2017年秋期講演大会(招待講演), 2017年.
4. 岡田真人, "スパースモデリングとエンジニアリングデータアナリティクス", Matlab Expo 2017 Japan(基調講演)(招待講演), 2017年.
5. Hikaru Takenaka, Kenji Nagata, Takashi Mizokawa, Masato Okada, "Bayesian Inference of Effective Classical Spin Hamiltonians from Hartree-Fock Calculation, American Physical Society March Meeting (国際学会), Baltimore, USA 2016/03/17.
6. 岡田真人, "スパースモデリングによる放射光データからの潜在的構造情報の抽出", KEK 放射光ワークショップ(招待講演), 2016年.
7. 岡田真人, "TIA の新展開とデータ駆動科学", 第8回 TIA シンポジウム(招待講演) 2016年.
8. 岡田真人, "AI for Science と脳型人工知能", 第19回情報論的学習理論ワークショップ(IBIS2016)(招待講演), 2016年.
9. Masato Okada, "Sparse modeling in condensed matter physics", 45th CEMS Colloquium, RIKEN(招待講演), 2016年.
10. Yasuhiko Igarashi, Kenji Nagata, Tatsu Kuwatani, Toshiaki Omori, Yoshinori Nakanishi-Ohno, Masato Okada, "Three levels of data-driven science", International Meeting on "High-Dimensional Data Driven Science" (HD3-2015)(招待講演)(国際学会), Mielparque Kyoto, Kyoto 2015/12/14.
11. Masato Okada, "Three levels of data-driven science", HD3-2015(招待講演)(国際

学会), Mielparque Kyoto, Japan, 2015/12/14 ~ 2015/12/14.

12. 岡田真人, "大規模計算とスパースモデリング", 海洋地球インフォマティクス 2015(招待講演), 海洋研究開発機構, 2015/05/13 ~ 2015/05/13.
13. 岡田真人, "レプリカ交換法を用いた変数選択", 第9回日本統計学会春季集会(招待講演), 明治大学中野キャンパス(東京都中野区) 2015年03月08日 ~ 2015年03月08日.
14. 岡田真人, "スパースモデリングとデータ駆動科学", 人工知能学会全国大会 2014(招待講演), ひめぎんホール(愛媛県松山市) 2014年05月12日 ~ 2014年05月15日.
15. Masato Okada, "Bayesian spectral deconvolution and sparse modeling", Cooperation of Computational Materials Science and Mathematics toward Smart Materials Design II - What Materials Informatics brings in? (招待講演), 東北大学 宮城県 2014年01月08日 ~ 2014年01月08日.
16. 岡田真人, "スパースモデリングの深化と高次元データ駆動型科学の創成 - 地球科学と脳科学のデータの背後に潜む普遍性 -", 資源・素材 2013(招待講演), 北海道大学, 北海道, 2013年09月03日 ~ 2013年09月03日.

[図書](計 0 件)

[産業財産権]

出願状況(計 0 件)

取得状況(計 0 件)

[その他]

1. 【領域主催公開シンポジウム】開催計 6 回(2014/12, 2016/03, 2016/12, 2017/06, 2017/12, 2018/03)
2. 【領域主催国際会議】開催計 2 回(2015/12, 2017/09)
3. 【新聞】海洋機構、地下断層のすべり現象-豊後水道で領域特定、地殻変動データ抽出, 日刊工業新聞, 2017, 7, 24
4. 【雑誌】「新リーダーの条件 日本を元気にする逸材 125 人」: 岡田真人, 文藝春秋, 2016 年 2 月号

5. 【雑誌】特集「使える！数学」：世界を変えた7つの方程式 週刊ダイヤモンド, 2016.1.23
6. 【公募説明会】開催計25回(2013年度14箇所500名以上参加, 2015年度11箇所400名以上参加)
7. 【新聞】ビッグデータの絞り込み より高速・高精度に：東大独自アルゴリズム活用, 日刊工業新聞, 2015.8.31
8. 【テレビ】NHKサイエンスZERO：情報科学の名探偵！魔法の数式スパーズモデリング, NHK Eテレ, 2015.8.23
9. 【新聞】サンデーサイエンス：宇宙や津波, 数学で迫る 少ないデータで「本質」解析, 日本経済新聞, 2015.5.3.
10. 【新聞】東京ドームに東大研究室 子供と科学の触れ合う場(TenQ), 日本経済新聞, 2014.8.25

## 6. 研究組織

### (1)研究代表者

岡田 真人 (Okada, Masato)

東京大学・大学院新領域創成科学研究科・教授

研究者番号：90233345

### (2)研究分担者

富樫 かおり (Togashi, Kaori)

京都大学・大学院医学研究科・教授

研究者番号：90135484

木川 隆則 (Kigawa, Takanori)

国立研究開発法人理化学研究所・生命システム研究センター・チームリーダー

研究者番号：20270598

谷藤 学 (Tanifuji, Manabu)

国立研究開発法人理化学研究所・脳科学総合研究センター・チームリーダー

研究者番号：60197530

駒井 武 (Komai, Takeshi)

東北大学・環境科学研究科・教授

研究者番号：30357024

宮本 英昭 (Miyamoto, Hideaki)

東京大学・大学院工学系研究科・教授

研究者番号：00312992

本間 希樹 (Honma, Mareki)

国立天文台・水沢 VLBI 観測所・教授

研究者番号：20332166

田中 利幸 (Tanaka, Toshiyuki)

京都大学・情報学研究所・教授

研究者番号：10254153

福島 孝治 (Hukushima, Koji)

東京大学・大学院総合文化研究科・教授

研究者番号：80282606

赤穂 昭太郎 (Akoo, Shotaro)

国立研究開発法人産業技術総合研究所・人間情報研究部門・研究グループ長

研究者番号：40356340

福水 健次 (Fukumizu, Kenji)

統計数理研究所・数理・推論研究系・教授

研究者番号：60311362

樺島 祥介 (Kabashima, Shosuke)

東京工業大学・情報理工学院・教授

研究者番号：80260652

藤代 一成 (Fujishiro, Issei)

慶應義塾大学・理工学部・教授

研究者番号：00181347

### (3)連携研究者

大森 敏明 (Omori, Toshiaki)

神戸大学・工学系研究科・准教授

研究者番号：10391898

永田 賢二 (Nagata, Kenji)

国立研究開発法人産業技術総合研究所・人工知能研究センター・主任研究員

研究者番号：10556062

桑谷 立 (Kuwatani, Tatsu)

国立研究開発法人海洋研究開発機構・地球内部物質循環研究分野・研究員

研究者番号：60646785