

戦略的創造研究推進事業 CREST
研究領域「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」
研究課題「広域撮像探査観測のビッグデータ分析
による統計計算宇宙物理学」

研究終了報告書

研究期間 2014年10月～2020年3月

研究代表者：吉田 直紀
(東京大学国際高等研究所カブリ数物
連携宇宙研究機構、教授)

§ 1 研究実施の概要

(1) 実施概要

すばる望遠鏡の主焦点カメラ Hyper-Suprime-Cam(HSC) を用いた広域宇宙サーベイは 2014 年に開始され、これまでに 220 晩ほどの観測が行われた。これは全体の計画の 8 割ほどに相当し、2019 年度中には観測は完遂される見込みである。

我々はこの宇宙観測ビッグデータを用いて、(1) 超新星をはじめとする時間変動天体を検出する画像解析技術を開発した。また、(2) 検出した変動天体の中から、宇宙の距離測定に用いることのできる Ia 型超新星を自動分類する深層学習アプリケーションを開発した。さらに、(3) 背景銀河の形状を歪ませる重力レンズ現象(光の伝搬に対する一般相対論的効果)の観測データにスパースモデリングの手法を適用し、宇宙の大域的物質分布を再現した。これまで 5 年の研究期間にわたって大規模な理論シミュレーションデータベースを構築しており、それらを用いて観測データを統計解析し、物質密度などの主要な宇宙論パラメータを測定するための研究をすすめている。

データ基盤グループ(筑波大/慶応大)は大規模データ処理のためのワークフローを利用してパイプラインの並列化および効率化を行い、画像データ処理の高速化をすすめた。宇宙論グループ(東京大)は観測進捗に合わせて継続的に画像データの質を検証し、時間差分画像解析のためのデータを作成した。画像解析機械学習グループ(NTT)と統計解析グループ(統計数理研)は差分画像解析によって超新星を検出するため、様々な機械学習で用いられている Random Forest を応用し、また最新の手法である pAUC ブーストや畳み込み Neural Network などの複数アルゴリズムを組み合わせた天体検出分類器を開発し、観測データ即時解析を行った。2016 年から行われた計 52 晩の変動天体観測では、6 万 5 千個の候補の中から 1800 個以上の超新星を抽出した。このうち 58 個は赤方偏移が 1 以上、つまり、約 80 億光年より遠くにあることが明らかになった。これほど遠くの Ia 型超新星は、これまで主にハッブル宇宙望遠鏡が過去 10 年間に実行した観測で発見された 50 個弱が知られているのみである。すばる HSC により発見された遠方超新星に対して分光追観測を行い、正確な距離を決定することで、宇宙の膨張史を明らかにする研究が飛躍的にすすむ。

宇宙論グループはまた、スーパーコンピュータを用いて宇宙の構造形成進化の理論計算(重力多体シミュレーション)を行い、101 個の異なる理論モデルに対する世界最大級の宇宙理論データベースを構築した。さらに統計解析グループと協働して、宇宙の 3 次元物質密度場の統計量を高速で計算する「エミュレータ」を開発した。この研究では、データ基盤グループは STR-T 木構造を利用し、シミュレーションデータ分析を高速で行うツールを開発し、宇宙論グループを支援した。開発されたエミュレータはすばる HSC サーベイに携わる多くの天文学者に提供され、宇宙論の研究が大きく進展している。

(2) 顕著な成果

<優れた基礎研究としての成果>

1. 機械学習を用いた天体画像解析により 1800 個の超新星を検出

概要:すばる HSC サーベイの観測画像データから検出された時間変動天体を自動的に選出するアプリケーションを開発した(森井ら 2016 [下記代表的な論文 1])。最大級の望遠鏡を使用する HSC サーベイ特有の課題である、暗い(遠い)天体の検出のために、Random Forest、畳み込み Neural Network などの従来技術に加え、AUC ブースト法や pAUC 法を新たに開発・最適化し、複数アルゴリズムの合議の結果を候補天体として出力する高精度の判別器を作成した。2015 年 8 月の観測では一晩の観測で明るい 10 個の超新星を発見し、観測当日のうちに速報を出すことに成功した。また、2016 年 11 月から行われた計 52 晩の変動天体観測では、6 万 5 千個の候補天体の中から 1800 個以上の超新星を抽出した(安田ら 2019 文献リスト

6-1 ①-69)。

2. 深層学習を用いた超新星分類器により遠方超新星を自動分類

概要：多くのタイプの超新星の中でも Ia 型超新星と呼ばれるものは、正確な距離を測定でき、宇宙膨張の時間変化を明らかにすることができる。我々は HSC サーベイにより得られた時間変動天体候補(上記 1 参照)から、深層学習を用いて Ia 型超新星を高精度かつ即時に検出する方法を考案した(木村ら 2017 [下記代表的な論文 2])。天文学で従来用いられていた基づく方法では、超新星の光度の時間変化を表すモデルライトカーブに実観測をフィッティングするアプローチを採用しており、一つの天体に対して多数回の観測が必要であった。これに対し、本研究では各観測バンド(波長帯)1回の観測データから Ia 型超新星であるか否かを判定する新手法を考案した。これには Ia 型超新星の候補を観測の初期の段階で迅速に絞り込むことができるという、分光追観測のための大きな利点がある。本手法は 2016 年 11 月から 2017 年 5 月にかけての観測にも適用され、より詳細な分光追観測を遂行する Ia 型超新星候補の選定の重要な指標として採用された文献リスト(鈴木ら 2017 文献リスト 6-3 ①-26)。

3. 重力レンズ現象の統計解析により宇宙の物質密度場を再構築

概要：すばる HSC サーベイにより得られた 1200 万個の銀河画像とその形状楕円率から重力レンズ強度(密度場シア)を測定し、スパースモデリングの手法に物質密度場を再構築した。天球面上で 137 平方度にもおよぶ広域の物質分布が再現された。さらに深層学習を用いた新たなノイズ低減法も提案した。従来の再構築法では主に背景銀河の数密度のばらつきに伴う誤差が物質密度場に含まれていた。本研究では敵対的生成アルゴリズムを用いて機械に重力レンズ「ノイズ」場を学習させ、再構築した密度場からノイズを除去することに成功した(白崎, 吉田, 池田 2019 文献リスト 6-1 ②-5)。背景銀河を地球からの距離によって選別することで 3 次元トモグラフィも可能であり、スパースモデリングを応用することで奥行き方向の密度場変動の不定性も低減すべく研究を加速している。

< 科学技術イノベーションに大きく寄与する成果 >

1. 観測的宇宙論への機械学習の導入

概要：我が国の天文学および宇宙物理学コミュニティが主導し、5 年間にわたって遂行する宇宙観測プロジェクトに対し、観測データ解析にいち早く統計的手法や機械学習をとりいれて実データ解析を進めた。観測期間の前半期には、データ処理パイプライン最適化とともに、統計解析に必要なツール開発をすすめることができた。口径 8-10 メートル級の巨大望遠鏡を用いる深宇宙探索としては世界最先端のデータ解析技術を有しており、2019 年度末に HSC サーベイが終了しその全データを解析する際には、宇宙の膨張史や宇宙の構造形成成長率など主要な宇宙論パラメータの測定に大きく貢献できる。さらに、観測データ統計解析に必要な理論シミュレーションデータベースの構築にも機械学習を導入し、世界に先駆けて高精度の解析をおこなった。宇宙画像取得から統計解析結果出力まで一連の作業を行うシステムが構築された。

2. 科学ビッグデータ処理技術の開発と応用

概要：計算機 CPU に最適化した STR-R 木による空間結合殻カウント法を開発した。宇宙観測データの解析や、宇宙の構造形成理論シミュレーションのデータ解析を行う際の基本技術の一つとして、天体周辺に存在する物質密度を積算する空間結合殻カウントが挙げられる。この作業の効率化には空間結合の一括構築が必要であるため、我々は STR-R 木アルゴリズムを提案し、使用する計算機に最適化して実装した。理論シミュレーションデータを用いた実験の結果、提案手法は高性能 R 木に対して 26.8 倍の性能改善を示した。この計算コードを GitHub で公開し、国立天文台のスーパーコンピュータで動作可能にしている。空間結合殻カウントは近傍探索などを行うデータ解析で頻繁に用いられる手法であり、本研究を異分野

に応用するためコード整備や公開を行った。

3. 高速計算ツール「エミュレータ」の開発

概要:2014年から国内最大級のスーパーコンピューターを用いて, 101個の理論宇宙モデルに対して計300試行の数値シミュレーションを行い, 宇宙の物質分布の理論データベースを構築した。その出力に対し, ガウス過程回帰に基づく統計モデルを採用し, 全データをベイズ的に学習し, 6つの主要な宇宙論パラメータ空間内の任意のサンプリング点(宇宙モデル)に対して物質分布の統計量を高速で計算し出力する「エミュレータ」を開発した。高精度エミュレータはすでに10件以上の天文観測データ解析プロジェクトで活用されている。

この開発過程で蓄積した技術、すなわちパラメータサンプリングの効率化や主成分解析, ガウス過程, ノンパラメトリック平滑化の組み合わせによる複合的統計計算は, その応用範囲が宇宙の構造形成シミュレーションに限定されたものではない。多数のパラメータセットに対して大規模なスーパーコンピュータシミュレーションを実行することが必要な科学研究や実社会問題など適用対象は広く、本研究チームでも応用分野を広げていく。

<代表的な論文>

1. Mikio Morii, Shiro Ikeda, Nozomu Tominaga, Masaomi Tanaka, Tomoki Morokuma, Katsuhiko Ishiguro, Junji Yamato, Naonori Ueda, Naotaka Suzuki, Naoki Yasuda, and Naoki Yoshida,
"Machine-learning Selection of Optical Transients in Subaru/Hyper Suprime-Cam Survey",
Publications of the Astronomical Society of Japan, (2016) 68 id104
2. Akisato Kimura, Ichiro Takahashi, Masaomi Tanaka, Naoki Yasuda, Naonori Ueda and Naoki Yoshida,
"Single-epoch supernova classification with deep convolutional neural networks",
IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW), (2017), pp. 354-359
3. Takahiro Nishimichi, Masahiro Takada, Ryuichi Takahashi, Ken Osato, Masato Shirasaki, Taira Oogi, Hironao Miyatake, Masamune Oguri, Ryoma Murata, Yosuke Kobayashi, and Naoki Yoshida,
"Dark Quest I. Fast and Accurate Emulation of Halo Clustering Statistics and Its Application to Galaxy Clustering", eprint arXiv:1811.09504, The Astrophysical Journal, (2019) 884 id29

§2 研究実施体制

(1) 研究チームの体制について

① 「宇宙論」グループ

研究代表者: 吉田 直紀 (東京大学 国際高等研究所 カブリ数物連携宇宙研究機構, 教授)

研究項目

- 撮像データ取得, 差分画像作成
- 理論シミュレーション

② 「統計解析」グループ

主たる共同研究者: 池田 思朗 (統計数理研究所 数理・推論研究系, 教授)

研究項目

- ・超新星の自動抽出に関する画像解析ソフトウェアの開発

③ 「画像解析機械学習」グループ

主たる共同研究者: 上田 修功 (日本電信電話株式会社 NTT コミュニケーション科学基礎研究所 上田特別研究室長)

研究項目

- ・時間変動天体の自動検出および分類方法の開発

④ 「データ基盤」グループ

主たる共同研究者: 川島 英之 (慶応義塾大学 環境情報学部 准教授)

研究項目

- ・解析パイプラインシステムの設計
- ・天文データ処理システムの設計

(2) 国内外の研究者や産業界等との連携によるネットワーク形成の状況について

・ データ基盤研究会創設

天文観測データはデータサイズが大きく、また、パイプライン処理および画像処理の内容が多岐に渡る。このようなデータ解析を対象にしたデータ基盤技術について研究議論を行う学会は、国内には存在しなかったため、情報共有やデータ基盤技術の発展が困難であった。そこで新たにビッグデータ基盤研究会という草の根研究会を組織した。組織委員会は、鬼塚真(大阪大学教授)、油井誠(Treasure Data 社)、そして主たる共同研究者である川島英之(慶応大学准教授)である。この組織では企業からアカデミアまで、データ基盤技術に関する興味や課題を有する者を集め、議論を行っている。

・ ギリシャのデータ科学研究者、天文学者との協働

本 CREST の活動を通して、ギリシャ クレタ大の天文学者および FORTH のデータ科学者との交流を継続してきた。2018 年には Andreas Zesus 准教授を東大に招いて天文観測データ解析に関する研究議論を行い、シンポジウムで講演していただいた。また 2019 年には研究代表者吉田がクレタ大および FORTH を訪問して今後の継続的な研究協力を計画議論した。2020 年の CREST 国際シンポジウムにもギリシャ人研究者を招聘予定であり、今後も天文ビッグデータ解析の共同研究を進める。

・ JST ビッグデータさががけとの協働

統計解析グループは、「ビッグデータ」さががけのプロジェクト「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」(研究代表者 酒向重行)と超広視野可視光カメラ Tomoe-gozen の新たなデータ処理法に関する共同研究を行った。Tomoe-gozen を用いた広視野観測は 2018 年度から開始されており、実データ解析がすすめられている。

・ 日独仏 3 国 SPPEXA 共同プロジェクト遂行

2015 年度から 2018 年度まで SPPEXA プロジェクト傘下にて、日本側では JST 本 CREST を通して理論シミュレーションコード研究開発を行った。専任研究員を 1 名雇用し重力流体計算コード開発をすすめるとともに、国内外の研究会で成果発表を行った。研究代表者吉田は期間中にドイツ ハイデルベルグおよびフランス ストラスブールを訪問し、研究進捗報告を行った。開発されたシミュレーションコードは 2019 年に公開を予定しており、次世代エクサフロップス級スーパーコンピュータ上でのアプリとして整備をすすめている。

・ 東京大学 知の物理学研究センターの設立

東京大学理学系研究科では、機械学習、深層学習や人工知能、さらには関連する自然科学研究の重要性と近年の発展を動機として、2018年12月1日に知の物理学研究センターを創立した。研究代表者吉田はこのセンターの立ち上げおよび運営に関わるとともに、本CRESTの研究をさらに加速してすすめる体制を整えた。具体的には、機械学習などの最先端の手法を物理学の研究に応用するだけでなく、深層学習やAIの動作原理を物理学の手法によって明らかにし、説明可能なAI(Explainable AI)の構築を目指している。そのために宇宙観測データや理論シミュレーションなど、本CRESTであつかうデータや手法を活用している。