

「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」

H27 年度
実績報告書

平成 26 年度採択研究代表者

吉田直紀

東京大学 大学院理学系研究科／カブリ数物連携宇宙研究機構
教授

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

§ 1. 研究実施体制

(1)「宇宙論」グループ

- ① 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)
- ② 研究項目
 - ・ 撮像データ取得、可視化
 - ・ 理論シミュレーション
 - ・ 銀河形成シミュレーションコードの開発(SPPEXA)

(2)「統計解析」グループ

- ① 主たる共同研究者: 池田 思朗 (情報システム研究機構 統計数理研究所、准教授)
- ② 研究項目
 - ・ 宇宙論パラメータ推定法の開発

(3)「画像解析機械学習」グループ

- ① 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、上席特別研究員)
- ② 研究項目
 - ・ 時間変動天体の自動検出法の開発

(4)「データ基盤」グループ

- ① 主たる共同研究者: 川島 英之 (筑波大学 システム情報工学研究科、講師)

② 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. 研究実施の概要

本研究の目的は、すばる望遠鏡に搭載された広視野撮像装置 (Hyper Suprime-Cam: HSC) を用いた広域宇宙観測のデータを解析するアプリケーションを開発することである。画像ビッグデータ解析により宇宙の物質分布や遠方超新星の距離を測定し、宇宙進化の歴史を明らかにすることを目指す。観測計画は平成 26 年度に開始され、現在まで支障なく観測を続けており、データの質も当初予定のとおりのもので得られている。

宇宙論グループは、宇宙の構造形成の大規模理論シミュレーションを開始し、重力レンズ現象の解析のためのデータベースを構築した。1つのシミュレーションでは、差し渡し 32 億光年という大領域を設定し、その中で 80 億個の質量粒子の重力 N 体進化を追う。このシミュレーションを、宇宙論パラメータを変えて 84 個行い、銀河重力レンズの統計量を正確に求めた。パラメータ空間探索にはラテン超格子法やガウス過程など、統計学からの知見を活用して探索効率を高めた。今後もシミュレーション数を増やし、最終的には数百個のシミュレーションデータベースを構築する。また、宇宙論グループは理論計算を加速するため、SPPEXA プロジェクトの下、新たな流体計算コードの開発を開始した。流体力学に加えて非平衡化学反応と輻射輸送を組み込むため、コードの構成を解析し、エネルギー方程式の変更と化学種分率を格納するためのデータ構成変更を行った。

統計解析グループと画像解析機械学習グループは協働し、すばる望遠鏡観測画像データから超新星を自動的に検出する判別器を作成した。平成 28 年度後半に本格的に始まる HSC 超新星探査では、Ia 型超新星を迅速に発見することが主目的の一つである。特に、明るさが増大する段階の Ia 型超新星を即時検出することが期待されている。これまでの観測から、HSC を用いた一晩の観測で検出される超新星は 100 個にもなると見積もられており、候補天体のすべてを人間が迅速に目視によって判別することは時間的に不可能である。H27 年度は最新の機械学習の方法を複数使い、それらの結果を組み合わせることによって、迅速に精度よく候補天体を検出する判別機を作成し、実際の予備観測に適用した。開発した判別機により、差分画像の中から効果的に候補天体を絞り込み、数十個の超新星を検出することができた(図1)。機械学習の方法を用いた判別機の構成およびシステムの成果に関しては論文を提出し、改訂中である。

	Information	Template (20" x 20")	New (20" x 20")	Subtracted (20" x 20")
1	HSC-SN15ay r = 23.1 RA 21:09:10.32 Dec -01:01:11.2 (2015-08-19.507)			
2	HSC-SN15az r = 23.8 RA 21:10:41.47 Dec -01:40:52.0 (2015-08-19.507)			
3	HSC-SN15ba r = 23.9 RA 21:12:33.25 Dec -01:04:24.7 (2015-08-19.507)			

図 1 機械学習による判別器によって得られた超新星候補例 (The Astronomer's Telegram # 7927、2015)

統計解析グループは、物質密度場から宇宙パラメータを推定するための統計量の検討を行い、宇宙パラメータの推定方法に関する研究も始めた。一般に統計的な推定の問題では、サンプルの個数あるいは観測領域の体積が増えれば精度が上がることは知られている。一方で、推定の精度には限界があり、それは Fisher 情報行列によって記述される。我々は与えられたデータのみからノンパラメトリックな方法で Fisher 情報行列を推定する方法の検討を行っている。

データ基盤グループは観測データ解析パイプラインにワークフローシステム Pwrake (Parallel Workflow extension for RAKE)を組み込んだ。これまでのパイプライン処理ではファイルアクセスがボトルネックになっていたが、PwrakeとGfarmを組み合わせることでファイルアクセスを並列化したところ、処理速度を改善することができた。次に、高速データベースの構築にも取り組んだ。具体的には近傍検索の空間索引と分散処理による高速化、再帰的インクリメンタル計算による集約処理の高速化、RDMA を利活用した MapReduce 処理系の開発を行った。