

「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」

H28 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、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、上田特別研究室長(NTT フェロー)、理化学研究所 革新知能統合研究センター、副センター長)
- 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、准教授)

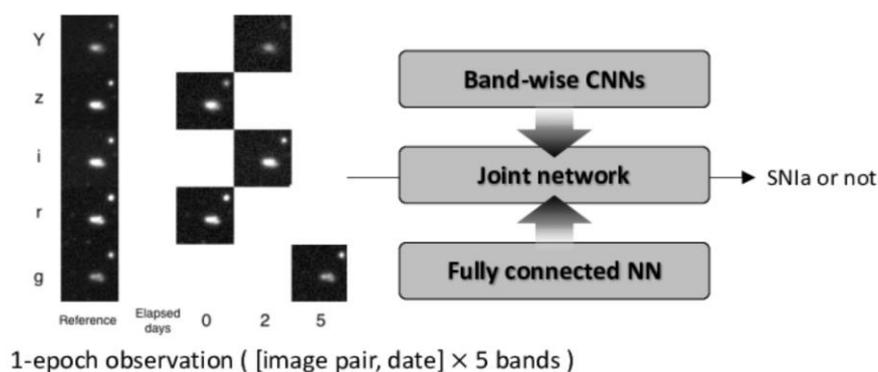
② 研究項目

- 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画像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多数の超新星や銀河を検出し、宇宙の物質分布や遠方超新星の距離を測定する。その結果を大規模シミュレーションデータベースと統合して統計解析し、宇宙の進化を記述する基本量(宇宙論パラメータ)を推定することを目指す。広域宇宙観測計画はH26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。H28年度は79晩の観測を実施し、全体で153晩の観測が終了した。またH28年度には、時間変動天体探索のための大型観測プログラムがおよそ6ヶ月にわたって遂行された。

宇宙論グループは、H27年度に開始した宇宙の構造形成の大規模理論シミュレーションを継続し、データベースを大幅に増強した。これらのシミュレーションは、本研究でH27年度までに作成した効率的な実験計画に基づき、100個の宇宙論パラメータセットに対して行われた。このデータベースを用いて、H27年度に実施した銀河重力レンズ統計量に対する機械学習フレームワークを拡張し、銀河の3次元空間統計量に対する学習を行った。



画像解析機械学習グループは、超新星の種類を分類する手法を開発した。今年度は、宇宙論で重要な役割を果たすIa型超新星の検出に注力し、突発天体候補の中から、Ia型超新星のみを自動的に検出する方法を考案した。上図のように天体の光度および観測日を特徴量として採用し、この特徴量を深層学習モデルに学習させ、実データを分類し、候補を挙げた。

統計解析グループは宇宙論グループと共同で、宇宙論パラメータを推定するための手法の検討を行った。また、さきがけ研究「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」(酒向重行)との共同研究により、動画データ圧縮法を開発した。行列分解を応用し、天文学的に重要な発見を逃すことなく1/10程度にデータ圧縮する方法を提案した。

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を組み込んだ。さらにPwrakeを用いた複数ジョブの並列実行に取り組んだ。高速トランザクション処理技法を開発しGfarmの高性能化に取り組んだ。次に、高速データベースの構築にも取り組んだ。まず、空間結合殻カウントを高性能化するために、CPUに最適化したSTR-R木を提案して高性能近傍探索を実現した。次に天体カタログデータ基盤の高性能化を検討すべく、shuffle処理をRDMAで高性能化した。