

研究課題別事後評価結果

1. 研究課題名： 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

2. 研究代表者名及び主たる研究参加者名（研究機関名・職名は研究参加期間終了時点）

研究代表者

吉田 直紀（東京大学大学院理学系研究科／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）

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）

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）

川島 英之（慶応義塾大学環境情報学部 准教授）

3. 事後評価結果

○評点：

A+ 非常に優れている

○総合評価コメント：

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（ダークマターの3次元分布）を再現した。このような超ビッグデータの大規模データ処理のための高速ファイルアクセス基盤技術の開発も同時に行い顕著な成果を得た。その結果、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が、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。過去10年間の観測でこれまで知られていた総数が50個弱であったことと比較すると、極めて大きな先駆的成果である。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、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」も開発した。このように、この分野で世界を先導し凌駕する極めて重要な高い研究成果を達成した。

本研究により、機械学習が観測的宇宙論の研究に先駆的に導入された。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することが期待できる。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解析が行われた。これらにより、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のパイプラインが構築された。また、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。この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国際的に活用されている。

このように、宇宙論という基礎科学の分野における未解決問題の解明の重要な糸口となるような世界的な先駆的成果を、ビッグデータ応用の観点から、機械学習の先進的研究成果を応用するのみならず、更に発展させることによって達成したことは極めて高く評価できる。