

## Z103a 機械学習と統計学を活用したビッグデータ宇宙論

吉田直紀 (東京大学カブリ IPMU)

2014年から開始されたすばる HSC サーベイは順調に宇宙観測画像データ取得をすすめており、今年度には初年度データリリースも予定されている。我々は2014年からCRESTプロジェクト「統計計算宇宙物理学」をすすめており、これまでに機械学習を用いた超新星検出や、大規模シミュレーションによる宇宙の物質分布の統計、観測データパイプラインの高速化、高速データベース構築など様々な研究を行ってきた。講演ではこのプロジェクトの取り組みと最近の成果を紹介する。超新星検出では Boosting 法、ランダムフォレスト、ディープネットをそれぞれ訓練し、それらの多数決によって判別を行う判別器を開発した。実際の画像に超新星の画像を埋め込んでデータの片寄りを補正し、AUC (Area Under the Curve) によって判別機を評価した。自動判別器を2016年度に行われる HSC 変動天体サーベイにも適用し、多数の Ia 型超新星を検出するための準備もすすめている。

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、重力レンズサーベイデータ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、1Gpc 立方の計算領域内で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100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し、共分散行列など宇宙論パラメータ推定のための統計量を計算する。講演の最後には、今後世界的に展開される大規模サーベイ計画とその巨大データ解析の展望を議論する。