

Y26a データサイエンス教育における天文データの活用

石橋史朗（会津大学）

近年、社会科学やビジネス分野においても、データサイエンスの重要性についての認識が高まっている。1つにはビッグデータやオープンデータなど、利用できるデータの範囲が急速に広がったこと、また一方で、データ処理技術自体の進展と普及がここ数年で進んだことが、その理由として挙げられる。コンピュータ理工学専門の会津大学においては、学生の多様な興味に対応して、様々なテーマを設定した「課外プロジェクト」と称する一連のコース群が開講されている。今回このコースの1つとして、天文データを利用したデータサイエンス講座を立ち上げたため、その取組み状況について報告する。

データ分析ツールとしては、入手が容易でかつ高機能を有する、フリーソフトの「R」を使用した。半期15コマの限られた時間ではあるが、基本統計量の計算から始めて、データマイニング的な分析手法まで進めることで、データサイエンスの基礎を習得すると同時に、今まで天文に馴染みのない学生に対して天文データにも興味をもってもらおうことを目指している。天文データは定量的データとともに、銀河の形態分類に代表されるような定性的データも多いため、社会一般で使われているデータサイエンス手法を学習する際のデータとしても、恰好の題材となっている。また実観測に基づく天体のカタログデータを扱うことにより、データ欠損の取扱いや、データ前処理の必要性など、データサイエンスの実践面での教育も可能となっている。講演では、授業の中で取扱った分析事例のいくつかについて具体的に紹介するとともに、データサイエンス教育における天文データ活用のメリットについても考察する。