

研究科重点プロジェクト

数学と諸分野の協働推進による
学際的・総合的な新領域研究の開拓

MATHEMATICS × EXTENSIVE SCIENCE

第1回講演会 兼 第48回応用数学連携フォーラム

日時

2016年12月8日(木)16時30分～17時30分

会場

東北大学 情報科学研究科棟 大講義室

講演者

池村 淑道(長浜バイオ大学 客員・名誉教授、
国立遺伝学研究所・総合研究大学院大学 名誉
教授)

タイトル

ゲノムビッグデータからの教師なし学習による
想定外知識発見と社会的重要な課題への応用

概要

ゲノム配列を代表例とする、生命科学分野の多様な測定データはビッグデータ化しており、そこにどの様な興味深い新知識が潜んでいるのかは想像することすら困難と思える。この様な状況下では、モデルや仮説や予備知識なしにビッグデータを研究できる「教師なし機械学習」が威力を発揮する。ゲノム配列はATGCの4種類の文字で構成される長文と言えるが、その文章中の連続文字列(例えば5連や6連、あるいは20連続文字)についての「ワードカウントに着目した機械学習」を行ってきた。ワードカウントのような理解しやすい解析を行うことでも、想定外の多様な知識発見が可能になる。20連文字を対象にした場合、約一兆一千億(=4²⁰)次元のデータ解析となり、スパースな大量データ解析となる。「コンピュータでエボラやインフルエンザウイルスの弱みを探る」等の社会的に重要な課題への応用も可能になっている。



<http://www.math.is.tohoku.ac.jp/~project/>