



統計的機械学習の理論構築とアルゴリズム開発

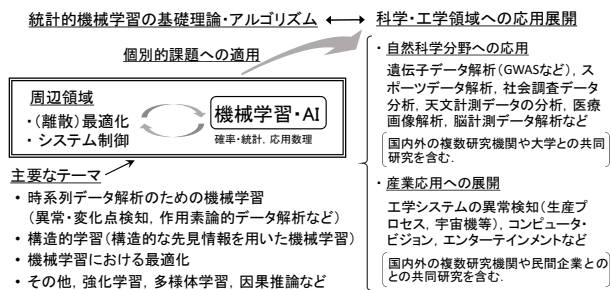
河原 吉伸

学位: 博士(工学)(東京大学)
 専門分野: 統計的機械学習、データ駆動科学

(統計的)機械学習は、多くの自然科学領域におけるデータ駆動による研究や、社会に浸透しつつあるAI関連技術を支える主要な基盤的研究分野の一つである。機械学習は元来、生物が持つ学習・認知能力の計算機上での実現を目標に発展してきた研究分野であり、統計科学や計算機科学などの幅広い数理科学分野を基礎とする複合領域である。近年では、高い予測能力の実現が可能な深層学習をはじめ、統計的機械学習がその研究の主流となっており、飛躍的な計測技術・情報インフラの発展を背景に急速に重要性が増している。本研究室では、

- (1) 統計的機械学習の新たな方法論の構築
- (2) 自然科学領域や産業分野におけるデータ解析への開発した方法・原理の応用

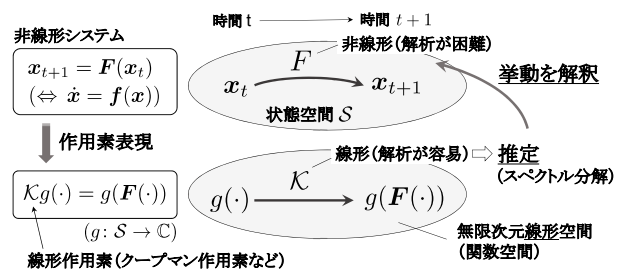
に関して、広く研究に取り組んでいる(図1も参照)。(1)に関しては、下記の2つの研究テーマについて昨今重点的に取り組んでいる。



(a) 多次元時系列データを用いた学習への作用素論的方法とその応用

時系列データを用いた学習や解析は、分野を問わず重要な課題である。特に数理的に扱いが困難となるのは、データを生成する機構が非線形性を有している場合である。作用素論的方法では、系の時間発展を関数空間における作用素表現で扱う事により、数理的に直接計算する事が困難である非線形性を回避して系を扱うアプローチである(図2)。この方法は、数理・応用物理分野における豊富な数理的枠組みに基づいた議論が行えると同時に、近年では、動的モード分解をはじめとした実用的な推定法が提案されており、広く科学分野で適用が進みつつある。本研究室では、作用素論的解析やその推定法、またその理論的解析の、機械学習的な数理的枠組みに基づく融合・拡張、または逆にこれら枠組みを

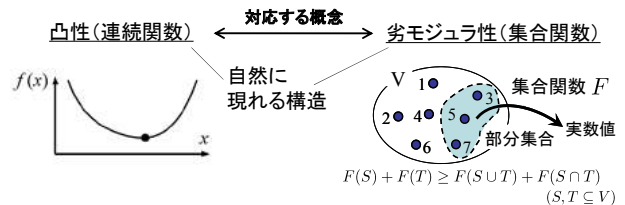
非線形現象の解析・予測へ利用するための方法論に関する研究に取り組んでいる。



また本研究テーマは、データ駆動による自然科学研究において重要な課題としても注目されている。本研究室においても、開発した方法を生命科学や脳科学などへ適用し、科学的知見獲得のための応用的研究も進めている。

(b) 機械学習における組合せ最適化

機械学習における問題解決場面では、組合せ的計算が本質的に重要となる場合が多い。例えば、予測器の構成において、可能な特徴(説明変数)の中から予測に重要となる特徴をいかに選択するかは、その性能に大きく関わる重要な課題である(特徴選択)。本研究室では特に、連続関数における凸性に対応する離散構造である劣モジュラ性に着目し、機械学習における組合せ的計算に関わる様々な課題に取り組んでいる。



その一つの重要な応用として、データの変数に関する構造的な事前情報(変数間のグループ関係やネットワーク状の依存関係など)を利用した学習が挙げられる(構造的学習)。劣モジュラ性は、このような学習の代表的な定式化である構造的スパース推定と深く関係していることが知られており、汎用的な手法開発や効率的な最適化の手がかりとなる。本研究室では、これに関連した新たな方法論の開発や理論解析などの研究にも継続的に取り組んでいる。