



統計的機械学習の理論と方法、及びその応用

河原 吉伸

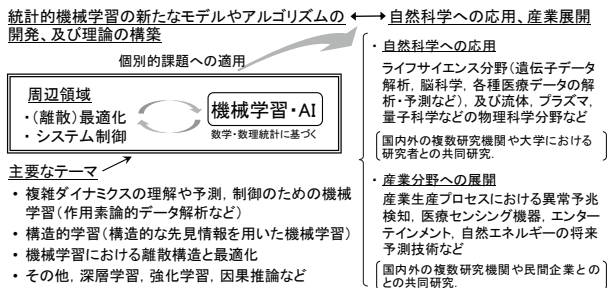
学位: 博士(工学)(東京大学)
専門分野: 統計的機械学習、データ科学

統計的機械学習は、様々な科学分野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関連技術を支える基盤的研究分野の一つです。機械学習は、元来は生物が持つ学習・認知能力の計算機上での実現を目的に研究が進められてきた分野で、幅広い数学・数理科学を基礎として発展してきました。近年では、高い予測性能を実現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ますます重要性を増しています。

本研究室では、下記のテーマを中心に、関連した課題へ幅広く研究に取り組んでいます(下図も参照)。

- (1) 統計的機械学習の理論と方法(新しい問題の発見と数理モデル化、アルゴリズムの開発、理論的解析)
- (2) 自然科学や産業分野における課題への開発した方法・原理の応用

(1)に関しては、最近では、下記の2つのテーマについて特に重点的に取り組んでいます。

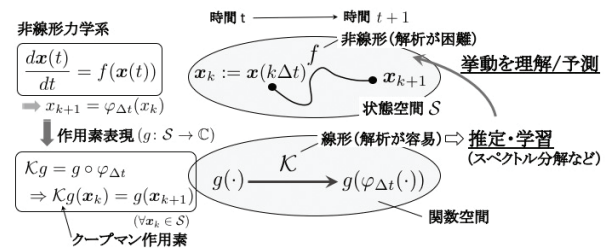


(a) 複雑なダイナミクスの理解や予測、制御のための機械学習

自然科学や工学において扱われる動的な現象や機構の多くは、非線形に起因した複雑な挙動を伴います。本研究室では、このような複雑なダイナミクスの理解や予測、そして制御に必要な、一連の統計的機械学習に関して研究を行っています。

例えば最近注力するテーマとして、非線形力学系の作用素論的解析に基づいた、機械学習やデータ解析の手法開発や理論構築を進めています(右上図)。作用素論的解析は、系の時間発展を関数空間における作用素表現で扱う事で、数理的に直接計算することが困難な非線形性を回避して力学系を扱うアプローチです。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いて、広く科学分野で適用が進みつつあります。

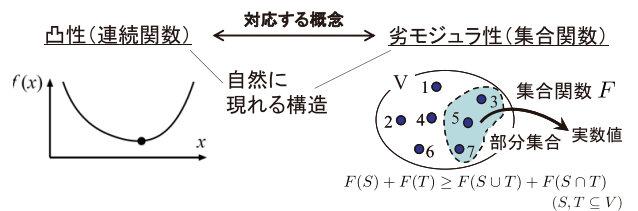


また本研究室では、様々な科学分野の研究者と協力して、それら分野における科学的知見の獲得を目的とした、開発方法の応用研究も積極的に進めています。

(b) 機械学習における離散構造と最適化

機械学習における問題解決場面や、学習モデル・アルゴリズムには、離散的な構造が本質的に重要となる場合が多く存在します。例えば、予測器の構成において、利用可能な特徴(変数、パラメータ)の中から予測に重要となるものをいかに選択するかという問題は、予測の性能に大きく関わる重要な課題として知られています(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の構造や最適化計算に関わる課題に取り組んでいます。



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ます(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いて、汎用的な手法開発や効率的な最適化の手がかりとなります。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ます。