

データ駆動科学の 情報数理基盤

東京大学 大学院新領域創成科学研究科
複雑理工学専攻
岡田真人

本スライドのまとめ

- 自然への理解を目指す科学には以下の二つのアプローチが存在する。
 - ミクロな基礎方程式からの演繹で、自然を理解できるという要素還元主義
 - ミクロな基礎方程式からの演繹だけでは、自然が理解できないという階層的自然観
- 本スライドでは、後者の階層的自然観の立場をとる。
- ある階層の数理的な記述に関して、データ駆動科学が必須であるという立場をとる。
- データ駆動科学を適用する際には、データを解析する情報数理基盤技術として、ベイズ推論とスパースモデリングの二つが必須であり、それらで十分であるという立場をとる。
- データの生成モデルが数理的に明らかでない場合、データの特徴量をスパースモデリングで選択する。
- データの生成モデルが数理モデルとして与えられている場合、ベイズ推論によって、データからモデルを選択する。また、数理モデルが複数ある場合、ベイズ推論を用いて、データからのモデル選択を行う。

データ駆動科学とは

- 機械学習などの人工知能を使い、各学問分野の問題を解いていく というアプローチ
- 実験/計測/計算データの背後にある潜在的構造の抽出に関して、データが対象とする学問に依存しない普遍的な学問体系
- 同じアルゴリズムがスケールや対象を超えて、有用であることが多いという経験的事実を背景として、その理由を問い、背後にある普遍性から、データ解析自体を学問的対象とする枠組み.

要素還元主義の破綻

- 要素還元主義：ミクロな基礎方程式からの演繹で、自然を理解できるという立場
- 要素還元主義はとても美しい立場であるが、これまでの科学の以下の発展の事例から考えて、要素還元主義は破綻していると言わざるを得ない。
- 事例1：量子力学と電磁気学で第一原理的な方程式が書き下せる物性物理学において、強相関電子系などの多対効果が無視できない系においては、計算機を使った第一原理計算でも、系の性質を記述することができないという事実がある。
- このような場合は、実験計測使って、系の性質を決めるパラメータをデータ駆動的に決めざるを得ない。
- 事例2：例えば恋愛感情のような高度な認知機能のメカニズムが量子力学や電磁気学から演繹できるとは誰も思わない。

階層的自然科学に対する 普遍的アプローチは存在するか？

- 階層的自然科学に対して、各階層個別にしか研究できないのか？各階層に依存せずに、研究することな可能な普遍的枠組みは存在するか？
- 我々は、その普遍的アプローチがデータ駆動科学にもとづく、スパースモデリングとベイズ推論であるというパラダイムを提案する。
- 階層によらずに、実験計測データを説明する数理モデルが存在する場合としない場合が存在する。
 - データを説明する数理モデルが存在しない場合、スパースモデリングを用いて、データそのもの、もしくはデータの特徴量から、知りたい特性を予測する。その際、スパースモデリングで特徴量を刈り込むことにより、データの生成モデルを考察する際の参考にする。
 - データを説明するモデルが存在する場合、ベイズ推論を用いて、数理モデルのパラメータをデータから決定する。さらに、数理モデルが複数ある場合は、ベイズ推論を用いて、データだけから、モデル選択する。