

データ駆動科学と民間就職 サイバーフィジカルシステム (CPS)の観点から

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

- データ駆動科学と機械学習は研究の目的は以下のように異なる。
 - 機械学習: 近似等を用いた高速アルゴリズムの開発が主目的。どのようなデータを解析するかは、インターネット上のデータなどで、あまり特定のセンサー系を対象とするものではない。
 - データ駆動科学: 実験計測系などを具体的に特定して、そこから得られるデータの背後にある構造や知見を、現代的な機械学習アルゴリズムを用いて抽出する。対象はセンサー系と情報処理系が強くカップルしたサイバーフィジカルシステム(CPS)と考えられる。
- サイバーフィジカルシステム(CPS): 現実(フィジカル) の情報を、コンピュータによる仮想空間 (サイバー) に取り込み、コンピューティングパワーによる分析を行った上でそれをフィードバックし、現実の世界に最適な結果を導き出すという、サイバー空間とフィジカル空間がより緊密に連携するシステム。

<https://monoist.itmedia.co.jp/mn/articles/2010/16/news059.html>

からの引用

- 民間企業では、機械学習のアルゴリズム開発をする事はまれであり、顧客や開発のニーズに応じたCPSを開発することが多い。
- つまりデータ駆動科学はCPSの観点からも、民間企業就職に有利な学問分野である。

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.

データ駆動科学におけるアルゴリズム

- データ駆動科学の数理情報基盤はスパースモデリングとベイズ推論。
- データの生成モデルが数理的に与えられていない場合はスパースモデリングを用いる。生成モデルが数理的に与えられている場合はベイズ推論を用いる。
- 得られる結果が近似アルゴリズムの数だけ存在するといふ状況でも考えられるので、スパースモデリングとベイズ推論のアルゴリズムとしては、計算量の削減のために提案された近似アルゴリズムを用いない。
- スパースモデリングの場合は全状態探索を用い、ベイズ推論ではMCMC法などの数値的厳密解法を用いる。

データ駆動科学への参入の容易さ

- データ駆動科学では、近似アルゴリズムを用いない方が望ましいので、近似アルゴリズムを勉強する必要はない。
- 具体的には、スパースモデリングの全状態探索とベイズ推論のMCMC法により数値的厳密解法の勉強をすれば良い。
- これは機械学習に比べて、データ駆動科学への参入の障壁が低いことを意味する。