

烏山・稲津研究室 研究室紹介

— 機械学習の基礎研究とデータサイエンスの実践 —



烏山 昌幸

KARASUYAMA, masayuki

Mail: karasuyama@nitech.ac.jp

おさらい: 近代的な AI・知的システムの核となるのが機械学習

人工知能

Artificial Intelligence

計算機上での知能の実現
経験から知識や規則を獲得

(統計的) 機械学習
(Statistical)
Machine Learning



統計学

Statistics

統計量の振る舞い・仮説検定
観測値から母集団の性質を推定



興味の対象から
有限個のデータ
(サンプル)を観測



深層学習
行列分解
カーネル法
マイニング
次元圧縮

ベイズモデル
スパースモデリング
最大マージンモデル
クラスタリング
表現学習

対象に関する知見
(モデル)



予測, 発見, 解釈
分解, 可視化

- Hand-crafted から **Data-driven** へ: 様々な分野の認識システム (画像, 音声など), 翻訳, 顧客情報解析, 遺伝子解析, 個別化医療, ボードゲーム, 結晶構造解析, etc.
- 多くの企業が機械学習に多大な投資: Google, facebook, Amazon, Yahoo!, Microsoft, Apple, IBM, Goldman Sachs, 楽天, CyberAgent, DeNA, etc.

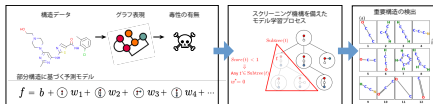
研究テーマ

AI 関連技術の汎用的な土台となる **機械学習基礎研究** とその **実データ解析応用**

基礎研究: 新しいデータ分析アルゴリズムの開発・効率化

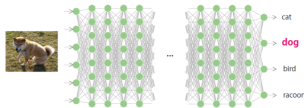
スパースモデルによる構造データからの知識発見

- 構造データ: DNA 配列, 化合物, 結晶構造など
 - ガンの subtype を左右する塩基配列は?
 - 化合物の毒性に影響する部分構造は?
 - 材料の強度を左右する結晶配置は?
- スパースモデル + マイニングで組み合わせ爆発を抑制:



ディープネットワークモデル (深層学習) の自動チューニングの研究 (AutoML)

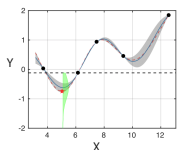
- 深層学習含む機械学習モデルは有用だがチューニングが大変!
- ユーザの指向に合うものをデザインするには? 複雑なデータ解析プロセスを最適化するには?



ベイズモデルによる未知空間の最適化・探索アルゴリズムの研究 (ベイズ最適化)

- X から出力 Y を予測する「順問題」ではなく、**所望の出力 Y を得る入力 X を探す「逆問題」**のための機械学習
- 例: 燃費 Y のよいモーターの設計 X は?, 伝導度 Y の高い結晶構造 X は?

機械学習の順方向(X→Y)モデル



所望のYを達成するXの確率的推薦

観測結果フィードバック

内部構造が陽に見えないシステム (ブラックボックス関数)

入力 X (実験設定, 合成候補物質, AIモデル)



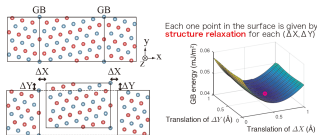
観測 Y (製品パフォーマンス, 物性値, モデル精度)

応用研究: 実世界のデータによる実践

材料科学のための機械学習

(マテリアルズインフォマティクス)

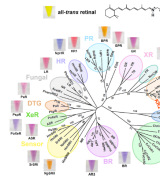
- 原子組成・構造からの物性予測
- 新規材料探索 (例: 固体電池材料)
- 物理シミュレーション・実験計測と機械学習の融合



時間コストの3%で発見
安定構造のAIシステム開発 名古屋工大
名古屋工業大学 未来材料・システム研究所
材料科学の分野で、材料の探索や設計に機械学習を応用する研究が進んでいます。この研究は、材料の安定構造を予測するためのAIシステムを開発し、材料の探索や設計に活用することを目的としています。このシステムは、材料の構造や組成から、その安定構造を予測し、材料の探索や設計に活用することができます。このシステムは、材料の探索や設計に活用するための重要なツールとなります。

生物データ解析のための機械学習

- アミノ酸配列・タンパク質立体構造からの機能予測
- 計測実験計画のための機械学習



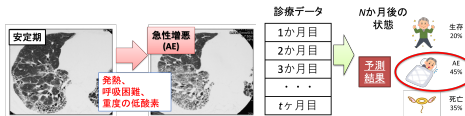
ロドプシンの
吸収波長予測
名古屋工業大学 未来材料・システム研究所
この研究は、ロドプシンの吸収波長を予測するための機械学習モデルを開発することを目的としています。このモデルは、ロドプシンのアミノ酸配列や立体構造から、その吸収波長を予測することができます。このモデルは、ロドプシンの探索や設計に活用するための重要なツールとなります。

<https://blogs.lt.vt.edu/stems/2014/05/01/optogenetics/>

診療データ (医学データ) による

リスク評価モデル構築のための機械学習

- 診療情報から N ヶ月後の状態を予測 (生存?容態悪化?死亡?)



応用研究は学内外の各ドメイン研究者と連携

- 材料系: 名工大 物理工学科, 名工大 生命・応用化学科, 名古屋大学 未来材料・システム研究所, 生物・医学系: 名古屋大学 医学部, 東京大学 物性研究所, 企業: 日本ガイシ, など

その他のこと

- 機械学習の基礎知識は勉強会などできちんと学びます
 - 得た知識を活かし、**データサイエンティスト**として就職する方も最近では多数いらっしゃいます
- 研究テーマは最初に大まかな希望を聞いて割り振りを決めます
- (参考) 数理的に雰囲気に近い講義
 - 数理情報概論, 機械学習論, データサイエンス, 知能プログラミング演習Ⅰの後半
- 目標は国際学会・国際論文誌発表！
 - 機械学習・データマイニングの有名 Top 会議 (ICML, KDD など) で発表した先輩もいます (Google など世界中の名だたる研究機関から論文が集まります)
- 「**主体的な**」研究への取り組みを期待します
 - (卒業のためだけにやるのではなく) **自発的に興味を持って積極的に取り組める研究室を選びましょう** 💡