令和5年度 Beyond 5G時代に向けた
新ビジネス戦略セミナー（第17回）

2023年9月1日

生成AIがもたらす社会変化と今後必要になる取組

理化学研究所／東京大学
杉山 将

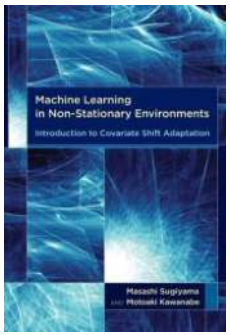
自己紹介

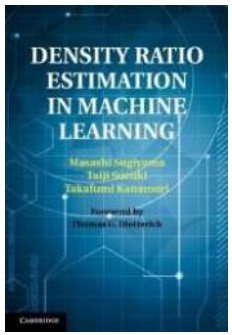
■現職：

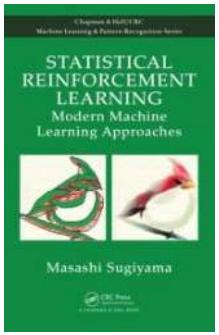
- 理化学研究所・センター長：研究者とともに
- 東京大学・教授：学生とともに
- 企業・技術顧問：経営者・エンジニアとともに

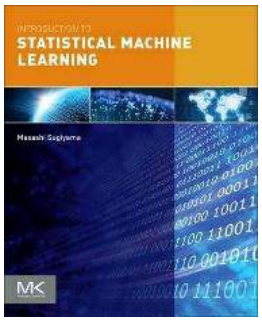
■専門分野：

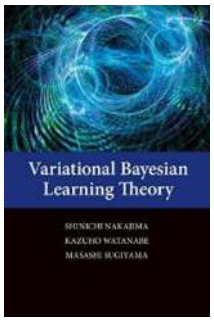
- 機械学習の数学的な基礎研究（コンピュータ，統計学など）
- 機械学習技術の実世界応用（画像，言語，音声，脳波，ロボット，医療，生命など）

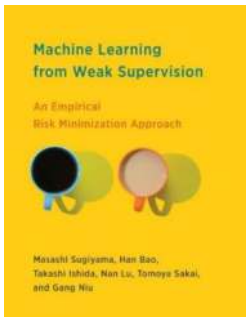












理研AIPセンター

研究体制:

- 汎用基盤技術研究グループ(杉山)
- 目的指向基盤技術研究グループ(上田)
- 社会における人工知能研究グループ(橋田)

統計情報:

- 常勤研究員128名(外国人41%, 女性23%)
- 客員研究員237名, 学生100名
- 延べ146名の海外インターン生
- 40+企業と共同研究
- 40+海外大学・研究機関と連携

研究拠点

日本橋オフィス

AIPの3つの研究グループ

汎用基盤技術（12チーム）

- 不完全情報学習チーム — チームリーダー 杉山 将 (Ph.D.)
- 構造的学習チーム — チームリーダー 河原 吉伸 (Ph.D.)
- テンソル学習チーム — チームリーダー Qibin Zhao (Ph.D.)
- 離散解析的学習チーム — チームリーダー Quang Minh Ha (Ph.D.)
- 圧縮情報処理ユニット — ユニットリーダー 田邨井 靖生 (Ph.D.)
- 深層学習理論チーム — チームリーダー 鈴木 大慈 (Ph.D.)
- 計算論的学習理論チーム — チームリーダー 畑埜 晃平 (Ph.D.)
- 因果推論チーム — チームリーダー 清水 昌平 (Ph.D.)
- 近似ベイズ推論チーム — チームリーダー Mohammed Emtiyaz Khan (Ph.D.)
- 連続最適化チーム — チームリーダー 武田 朗子 (Ph.D.)
- 数理科学チーム — チームリーダー 坂内 健一 (D.Math.Sci.)
- 高次元因果解析チーム — チームリーダー 令泉 允聖 (Ph.D.)

- 学習と推論のアルゴリズム
  - ロバスト，ベイズ，構造，…
- 統計的学習と最適化の理論
  - 収束性，表現力，非凸，…

目的指向基盤技術（17チーム）

- がん探索医療研究チーム — チームリーダー 浜本 隆二 (Ph.D.)
- iPS細胞連携医学的リスク回避チーム — チームリーダー 上田 修功 (Ph.D.)
- 分子情報科学チーム — チームリーダー 津田 宏治 (Ph.D.)
- 認知行動支援技術チーム — チームリーダー 大武 美保子 (Ph.D.)
- 防災科学チーム — チームリーダー 上田 修功 (Ph.D.)
- インフラ管理ロボット技術チーム — チームリーダー 岡谷 貴之 (Ph.D.)
- 医用機械知能チーム — チームリーダー 原田 達也 (Ph.D.)
- データ駆動型生物医科学チーム — チームリーダー 竹内 一郎 (Ph.D.)
- 計算脳ダイナミクスチーム — チームリーダー 山下 由紀 (Ph.D.)
- 遺伝統計学チーム — チームリーダー 田宮 元 (Ph.D.)
- 病理情報学チーム — チームリーダー 山本 博一郎 (M.D., Ph.D.)
- 自然言語理解チーム — チームリーダー 乾 健太郎 (Ph.D.)
- 知識獲得チーム — チームリーダー 松本 裕治 (Ph.D.)
- 言語情報アクセス技術チーム — チームリーダー 関根 聡 (Ph.D.)
- 空間情報学チーム — チームリーダー 横矢 直和 (Ph.D.)
- 食薬情報知能チーム — チームリーダー 浜中 雅俊 (Ph.D.)
- 食糧情報理解チーム — チームリーダー 吉井 和佳 (Ph.D.)

社会におけるAI（6チーム）

- 社会におけるAI利活用と法制度チーム — チームリーダー 中川 裕志 (Ph.D)
- 科学技術と社会チーム — チームリーダー 佐倉 統 (Ph.D.)
- 分散型ビッグデータチーム — チームリーダー 横田 浩一 (Ph.D.)
- 経済経営情報融合分析チーム — チームリーダー 星野 崇宏 (Ph.D.)
- 人工知能セキュリティ・プライバシーチーム — チームリーダー 佐久間 淳 (Ph.D.)
- 人工知能安全性・信頼性ユニット — ユニットリーダー 荒井ひろみ (Ph.D.)

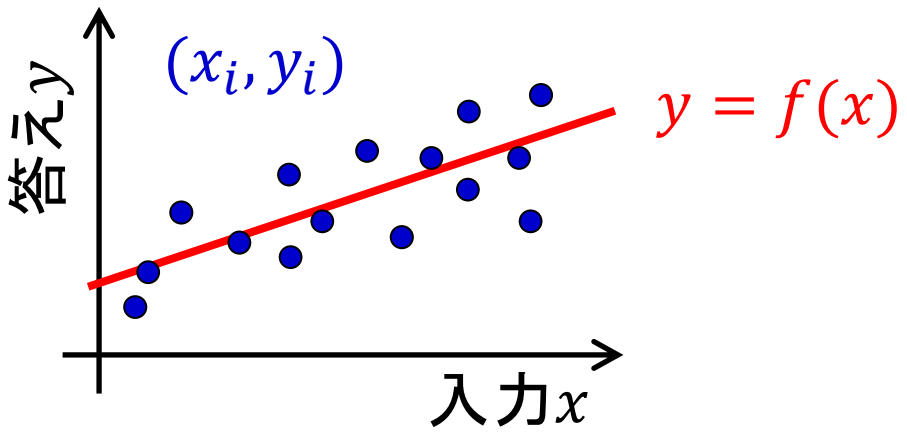
- 個人データ管理
- AIセキュリティ
- 経済と経営
- 倫理とガバナンス

- AIによる科学研究の加速
  - がん，材料，遺伝子，…
- AIによる社会課題解決への貢献
  - 自然災害，高齢者ヘルスケア，教育，…

通常の機械学習

■ 識別モデルの教師付き学習：

- 質問xに対する答えyを教え，
  入出力関係$y = f(x)$を学習



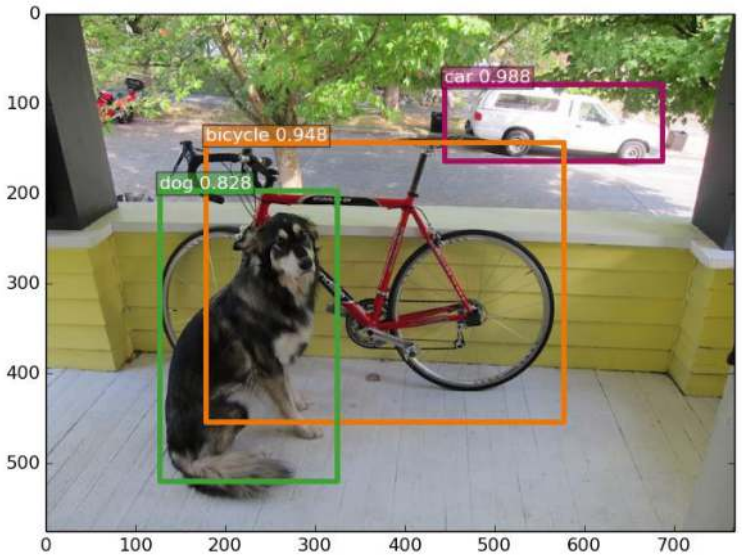
- 最小二乗法：$\min_g \sum_i (y_i - g(x_i))^2$
- 未知の入力x'に対する答えy'が予測できる

識別モデルの例

音声認識（声→言語）

https://spjai.com/speech-recognition/

画像理解（画像→物体）



https://github.com/zhreshold/mxnet-ssd

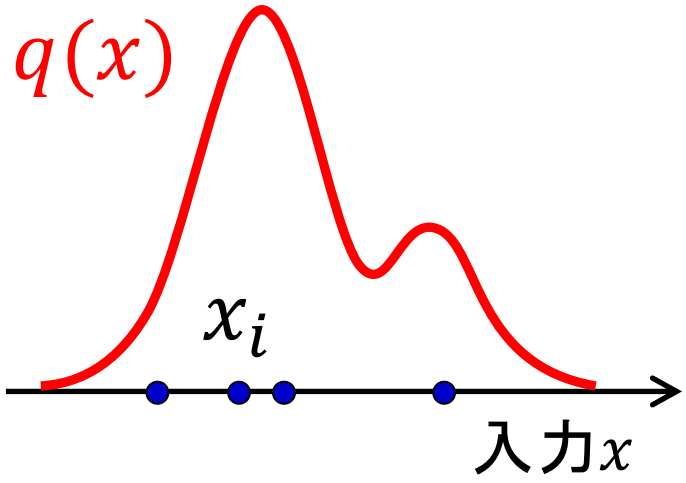
言語翻訳（日本語→英語）

https://toyokeizai.net/articles/-/167690

生成AI

■ 生成モデルの教師なし学習：

- データxの例から，データの生成確率$p(x)$を学習



- 最尤推定法：$\max_q \prod_i q(x_i)$
- 学習した確率分布もとに，新しいデータx'が生成できる

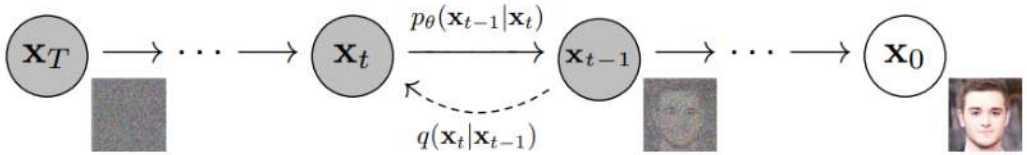
生成モデルの例:画像

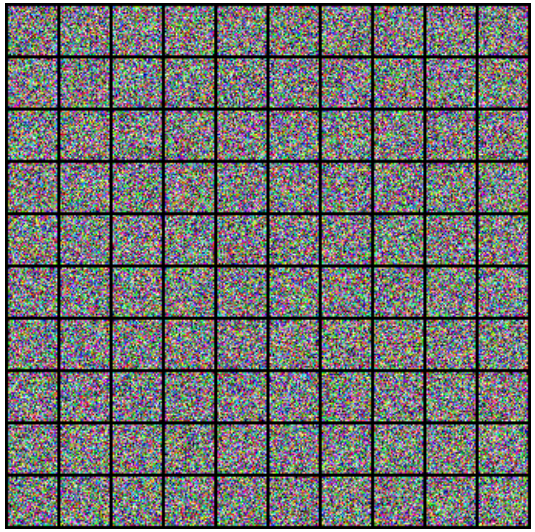
■拡散モデル

- 学習:画像に徐々に
  雑音を加えていく

- 生成:雑音から徐々に
  雑音を除去

https://arxiv.org/pdf/2006.11239.pdf

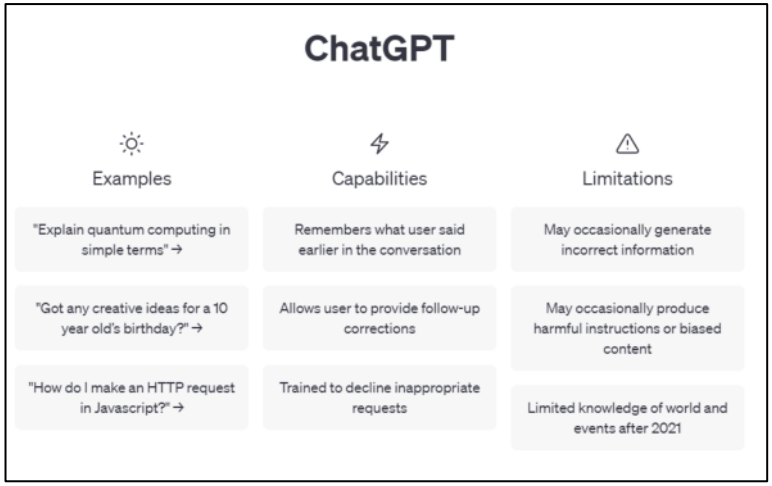




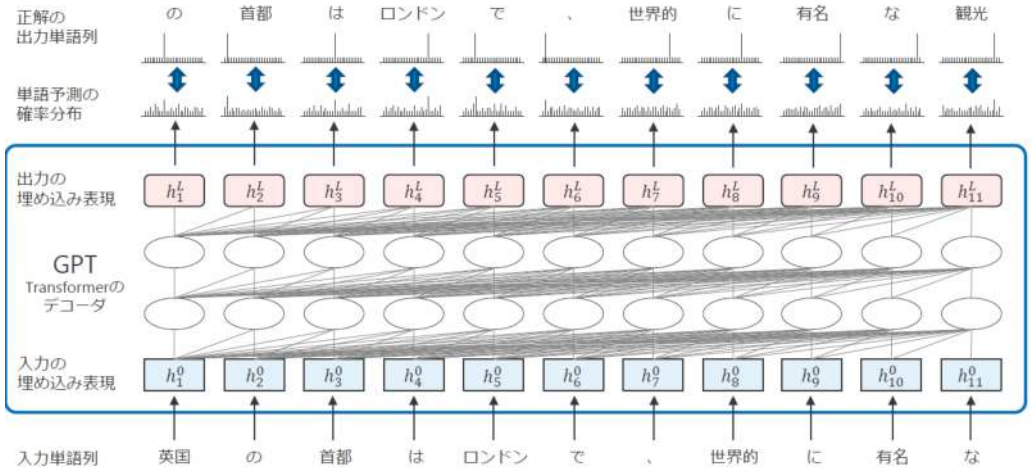
https://yang-song.net/blog/2021/score/

生成モデルの例：言語

- GPT (Generative Pre-trained Transformer)
  - 学習：文章の分布を学習
  - 生成：与えられた文章の次の単語を生成



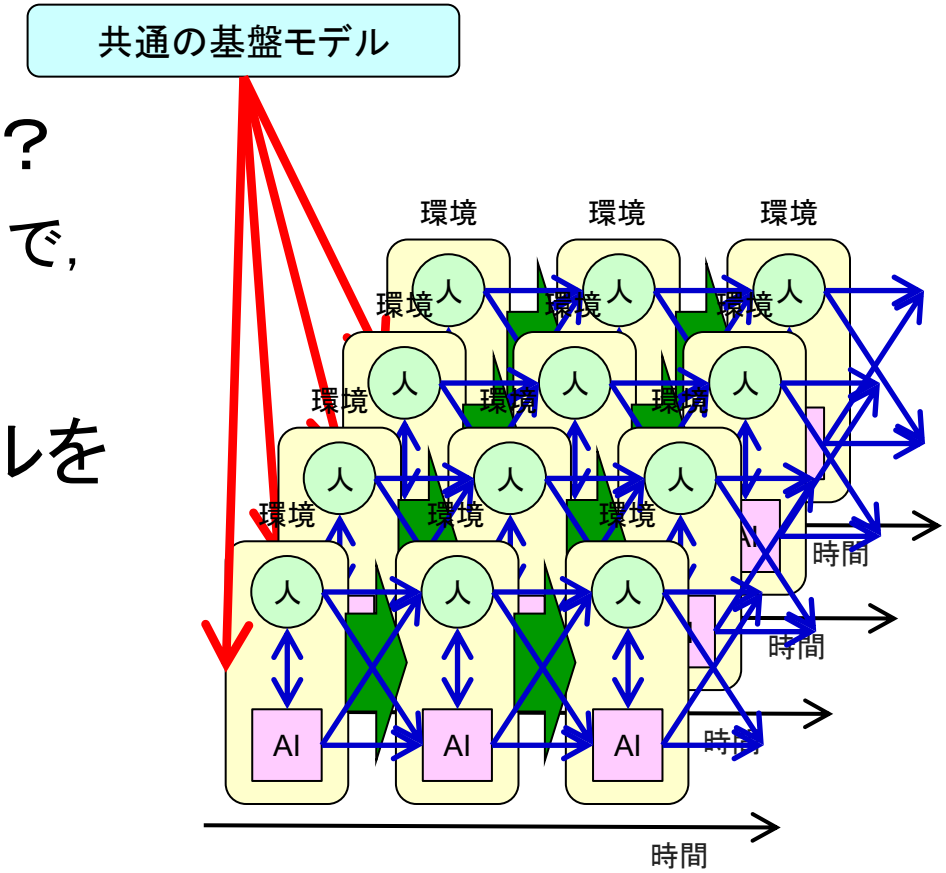
https://chat.openai.com/



https://speakerdeck.com/chokkan/20230327_riken_llm

生成AIの個別化

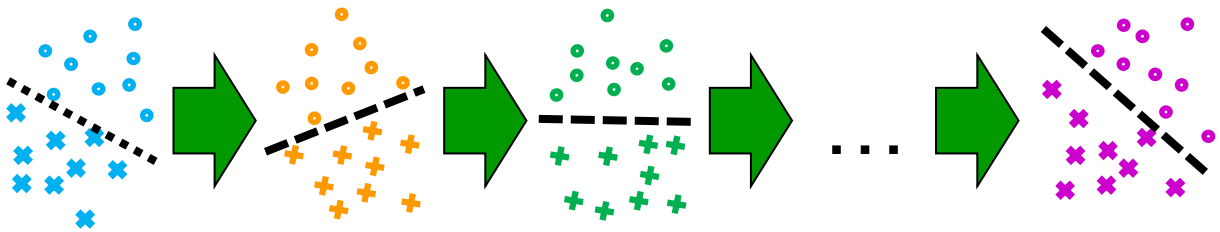
- 日本独自の生成AIは必要か？
  - Yes. 仮に技術的に新規性がなくとも，
    日本の文化を反映させた生成AIを普及させるべき
- 今の言語モデルは，
  事後確率最大の解を出力？
  - 多様な意見があるはずなので，
    事後分布を示すべき
- そもそも，1つの言語モデルを
  皆で使うのはおかしい？
  - 人に個性があるように，
    言語モデルも個別化すべき



重要な研究課題

■ オンライン学習：

- 基盤モデルに最新の情報を追加していく必要がある
- 巨大モデルの再学習は不可能．逐次学習が不可欠



■ 因果推論：

- 相関関係だけでなく，
  因果関係の分析が重要
- 介入した場合の確率の変化や，
  反実仮想的な確率を推測

ノーベル賞受賞者数

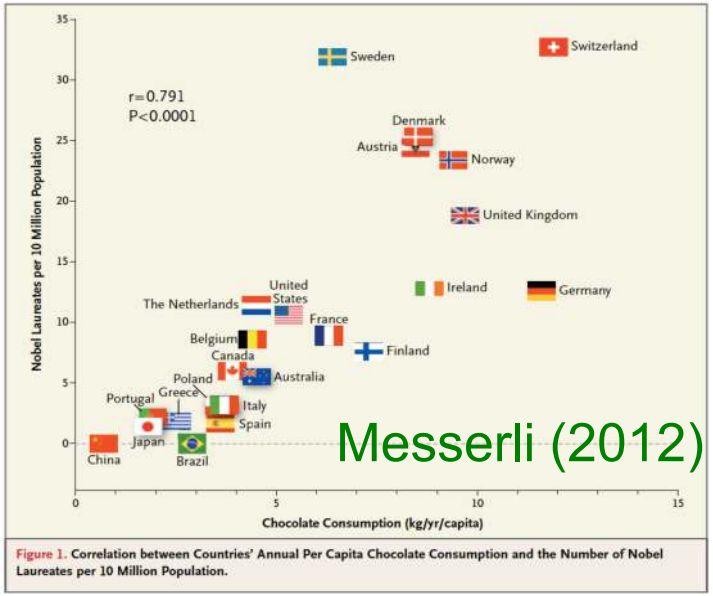


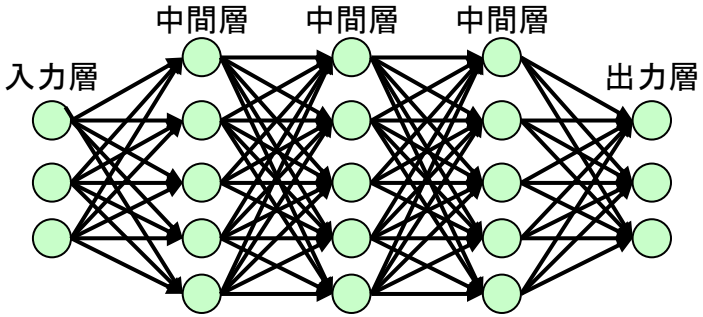
Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

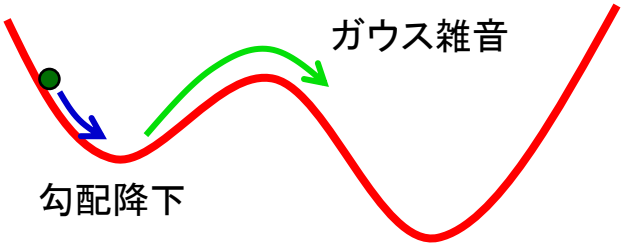
チョコレート消費量

重要な研究課題

■ 深層学習の理論:

- 深層学習はなぜ浅層学習より予測性能が高いのか？
- トランスフォーマーの汎化能力は？
- なぜ大域的最適化が可能なのか？



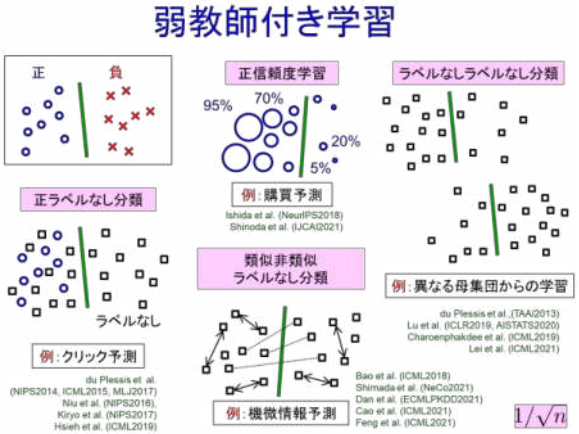


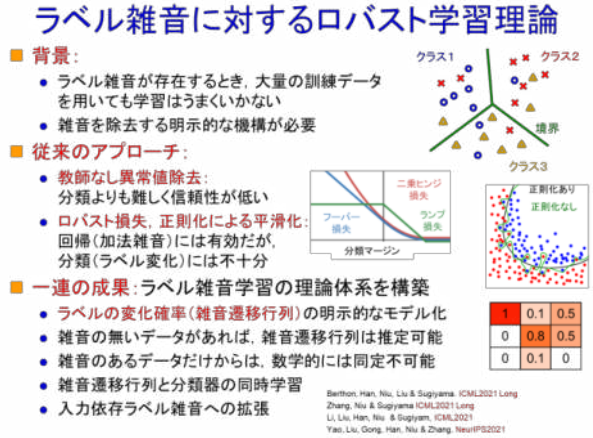
■ 弱教師付き学習:

- 限られた教師情報からの高精度学習

■ 雑音ロバスト学習:

- 教師情報に含まれる雑音を抑制しながら高精度に学習





日経新聞　経済教室（2023年3月30日）

メディアで「AIが学習する」という言葉を頻繁に見聞きする。擬人化されたAIが私たちの思いをくみ取って自発的に学習してくれるわけではなく、技術者がプログラムをつくってコンピューターを学習させて

画像生成やテキスト対話型の人工知能（AI）システムが注目を集めている。専門的な知識は必要とせず、簡単な言葉を入力するだけで実用的な出力が得られるため、AIの活用は新たな局面に突入したといえそうだ。本稿では、AIシステムの基本となる機械学習の技術的な概要と、その本質について述べたい。

機械学習とは、機械（コンピューター）に学習能力を持たせるための情報技術の総称である。音声や画像の認識、言語の翻訳、情報の検索、商品の推薦、顧客情報の分析、株価の予測、サイバー攻撃の検知、自動運転車の制御など、様々な場面で用いられている。

すぎやま・まさし　74年生まれ。東京工業大博士（工業）。専門は機械学習。理化学研究所センター長

AI開発の現在地 ㊤

杉山将 東京大学教授

原理解明さらなる研究必要

https://www.nikkei.com/article/DGXZQOCD152XM0V10C23A3000000/

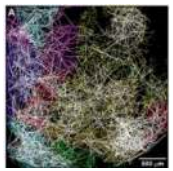
- さて、このような新AI時代にどう対応すべきだろうか。少なくとも、人間によるAI技術を活用する能力と、AIが生成した情報の真贋を見抜く力は必要だ。
- ただし真贋の判定はまさに機械学習における教師付き学習問題そのものであり、人間が真贋を判断するとは、結局はAIが自動生成した入力データに、人間が教師情報を付与するのと同じことである。
- いまはAI技術の現実的な発展のために身を粉にして教師情報作りを続けるべきか、それとも新たな知性の獲得のため原理の解明を目指すべきか。まさに議論が必要である。

理化学研究所における生成AI研究

DXを活用し、異分野の連携融合を促進する「つなぐ科学」

先端計算科学と予測アルゴリズム、データ整備を連環させ、
未来の予測制御の科学を開拓

良質なデータ整備



脳の神経活動の
リアルタイム
可視化技術(例)

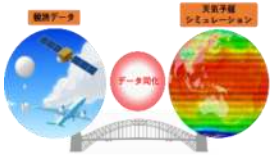
SPring-8
SACLA

バイオリソース

研究DXの先駆的取組へ発展

データ　AI　スパコン

AI×数理で
予測の科学を開拓



学習理論

AI

基盤モデル・生成AI

計算可能領域の拡張
量子古典ハイブリット

量子コンピュータ

スパコン「富岳」

狙い
研究サイクル
加速

「未来の予測制御の科学」を
分野の枠を超えて開拓

RIKEN