

2025/03/15 尾崎 克久

## 機械学習とは？

みなさんは「人工知能（AI）」という言葉聞いたことがありますか？スマートフォンの音声アシスタントや、ネットでおすすめの商品を紹介してくれるシステムなど、AIは私たちの身近なところで活躍しています。最近では、ChatGPT・Gemini・Claudeなど生成AIが話題です。そのAIを支えているのが「機械学習」という技術です。

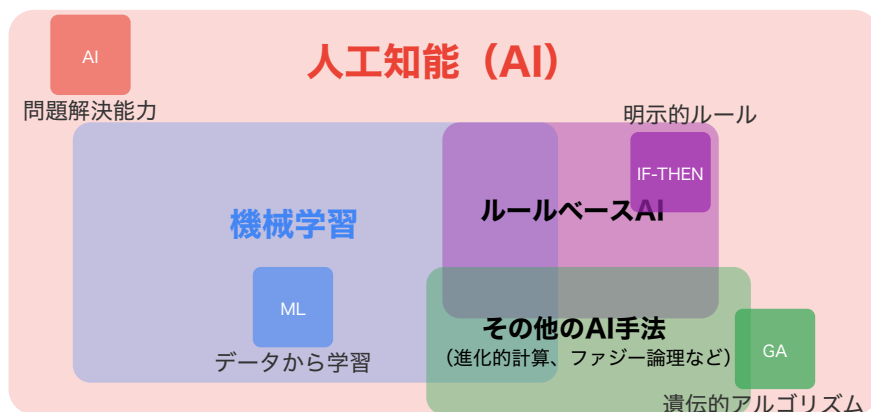
## 人工知能と機械学習の違い

人工知能（AI）とは、人間の知的な処理を模倣することを目的としたシステムや手法の総称です。ひとくちにAIと言っても、中にはさまざまな方法があります。その一つが「機械学習」です。

## 機械学習とは

機械学習とは、コンピュータが大量のデータを学び、その中にあるパターンを見つけ出す技術のことです。人間がすべてのルールを教えなくても、コンピュータ自身がデータをもとに学習し、新しい情報に対しても判断できるようになります

## 人工知能と機械学習の関係



AIは人間の知的処理を模倣するシステムの総称、機械学習はその主要技術の一つ

## 機械学習の基本的な仕組み

機械学習には、大きく分けて「教師あり学習」「教師なし学習」「強化学習」という3つの種類があります。

- 教師あり学習
  - 例えば、「この植物はアゲハチョウが好む」「この植物は好まない」といった正解データをコンピュータに与えて学習させます。
  - コンピュータは与えられたデータをもとに、新しい植物を見たときに「これはアゲハチョウが好きそうかどうか」を予測できるようになります。

- さらに、「植物の成分データからアゲハチョウが訪れる確率を予測する」といった回帰分析も教師あり学習の一例です。
- 教師なし学習**
  - 正解を教えずに、大量のデータの中からパターンや特徴を見つける方法です。
  - 例えば、アゲハチョウが産卵する植物の化学成分データを分析し、似た特徴を持つ植物を自動的に分類する（クラスタリング）。
  - こうすることで、人間が気づかなかった新しい関係性を発見できるかもしれません。
- 強化学習**
  - これは「試行錯誤しながら学ぶ」方法です。
  - 例えば、ゲームをプレイするAIは、最初は適当に動きますが、得点が高くなる動きを繰り返すことで、だんだん上手になります。
  - 強化学習では、「環境」「状態」「行動」「報酬」の4つの要素を考慮しながら最適な選択を学習していきます。

## 機械学習の3つの主要手法



## 生物学研究における機械学習の役割

生物学の研究では、大量のデータを扱うことが多くあります。私の研究に関わることで例を挙げると、植物が持つ化学成分のデータや、アゲハチョウの味覚神経の電気信号のデータなどです。

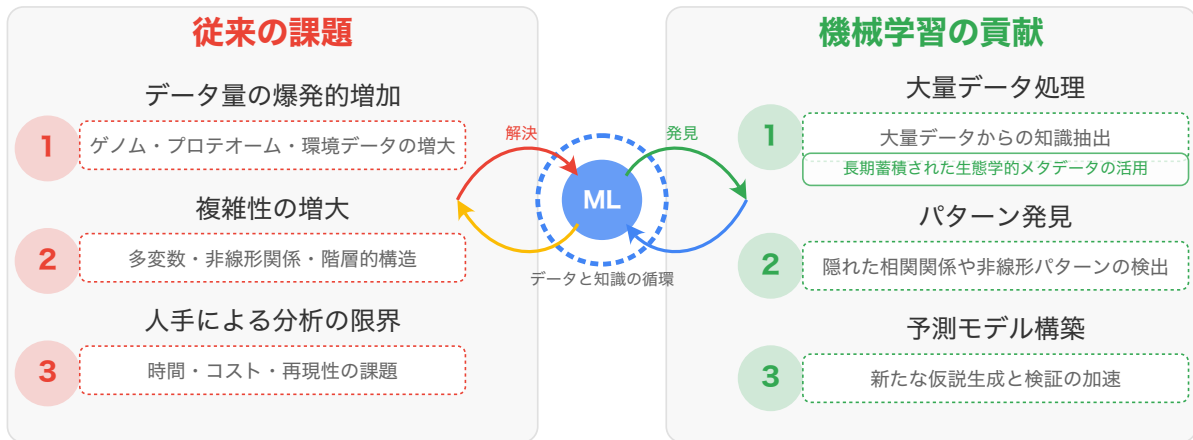
しかし、こうしたデータを人間が手作業で分類したり、関係性を見つけたりするのはとても大変です。そこで、機械学習を使うと、

- 植物の化学成分を分析し、どの成分がアゲハチョウの食草選択に影響を与えているのかを探ることができる。
- 味覚神経の電気信号を解析し、どんなパターンのときにアゲハチョウが産卵するのかを見つけることができる。

このように、機械学習は生物学の研究においてもとても役立つ技術なのです。

## データ駆動型生物学の実現

生物学研究における機械学習の役割



機械学習は生物学研究のパラダイムシフトを加速し、新たな発見への道を開いています

## 生物学研究における機械学習の幅広い応用

機械学習は様々な生物学の分野で革新的な研究手法として活用されています：

### 生態学・生物多様性研究

- 種の分布モデリング：気候や地形などの環境データから、生物種の分布域を予測します。
- 生物音響学：鳥や哺乳類の鳴き声を自動識別し、個体数や行動をモニタリングします。
- カメラトラップ画像の自動分析：野生動物の個体識別や行動パターンを機械学習で解析します。

### 構造生物学・分子生物学

- タンパク質-リガンド相互作用の予測：創薬研究に応用されています。
- RNA二次構造予測：非コードRNAの機能解析に役立ちます。
- 代謝経路予測：新規代謝経路の発見や合成生物学への応用が進んでいます。

### マイクロバイオーム研究

- 細菌群集構造の予測：環境要因からマイクロバイオーム組成を予測します。
- 宿主-微生物相互作用のモデリング：健康状態と腸内細菌叢の関連性を分析します。
- メタゲノム解析：環境サンプルからの新規微生物機能を予測します。

### 進化生物学

- 系統樹推定：大規模分子データからの系統関係を推定します。
- 進化速度予測：遺伝子の進化速度と機能的制約の関係を解析します。
- 祖先配列再構築：絶滅種の特性を推定することも可能になります。

### 作物科学・農業応用

- 作物の収量予測：気象データと遺伝情報から収量を予測します。
- 病害虫の発生予測：環境条件から病害虫の発生リスクを予測し、効率的な農業管理に役立っています。
- ゲノミックセレクション：次世代育種への応用が進んでいます。

## 神経科学

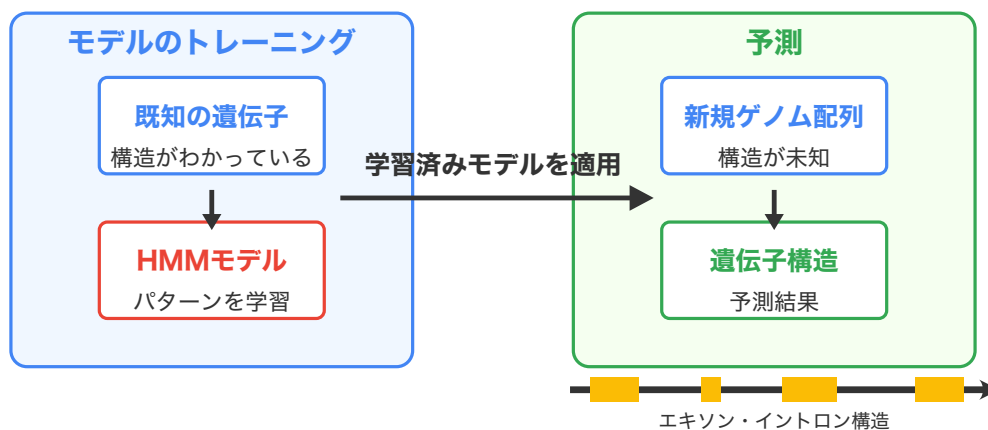
- **脳活動パターン解析**：fMRIデータからの認知状態を解釈します。
- **神経回路のマッピング**：コネクトーム解析により、神経回路の構造と機能を理解します。
- **行動予測**：神経活動データからの動物行動を予測します。

## 遺伝子予測ツールAugustusと機械学習

機械学習は、ゲノム解析の分野でも重要な役割を果たしています。例えば、遺伝子の構造を予測するツールとしてAugustusが広く使われています。

- **Augustusは機械学習の一種を利用**
  - Augustusは、**隠れマルコフモデル (HMM, Hidden Markov Model)** を用いた統計的手法で遺伝子の構造を予測します。
  - HMMは確率モデルを用いてデータの隠れたパターンを解析するため、従来の決定論的なルールベースの方法よりも柔軟な予測が可能です。
- **どのように遺伝子を予測するのか？**
  - 遺伝子には、「エキソン」「イントロン」「プロモーター」「終止コドン」などの構造があります。
  - Augustusは、既知のデータからこれらのパターンを学び、新しい配列に対して「この部分はエキソンである確率」「ここはイントロンである確率」を計算して予測を行います。
- **他の遺伝子予測ツールとの比較**
  - GeneMark もHMMを使って予測を行いますが、GeneMark.hmm-ES などのバージョンでは教師なし学習も利用します。
  - 近年では、**ディープラーニング (深層学習)** を活用した遺伝子予測ツールも開発されており、より複雑な特徴を学習できるようになっています。

## Augustusによる遺伝子予測



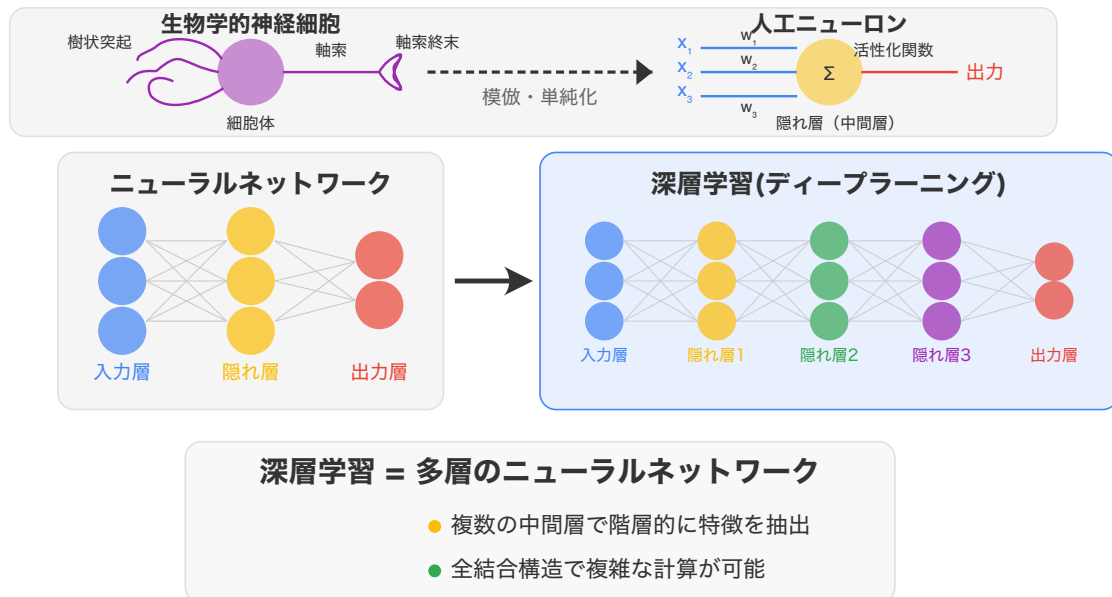
既知の遺伝子からパターンを学び、新しい配列の構造を予測

## 深層学習（ディープラーニング）について

深層学習は、機械学習の一分野であり、多層のニューラルネットワークを用いて、データから複雑な特徴を自動的に抽出する能力に優れています。従来の機械学習手法（HMMなど）と比較して、

- **より複雑な特徴の自動抽出:** 人間が手作業で特徴量を設計する必要がなく、データから自動的に学習します。
- **大量のデータからの高精度なモデル構築:** 大量のデータを扱うことで、より複雑なパターンを捉え、高精度な予測モデルを構築できます。

## ニューラルネットワークと深層学習



生物学研究における深層学習には、以下のような主要なアーキテクチャが応用されています：

- **畳み込みニューラルネットワーク (CNN)**
  - ゲノム配列上の特定のパターン（モチーフ）を検出するのに優れています。
  - 例：DeepBindは転写因子結合部位を高精度に予測し、DeepSEAはDNAの非コード領域の機能を予測します。
- **リカレントニューラルネットワーク (RNN)**
  - 配列データの長距離依存性を捉えるのに適しています。
  - 例：LSTMを用いたDEEP-TARGETは、miRNAのターゲット遺伝子を予測します。
- **Transformer**
  - 自己注意機構により、配列内の任意の位置間の関連性を効率的に学習できます。
  - 例：AlphaFoldはタンパク質構造予測で革命を起こし、ESM (Evolutionary Scale Modeling) はタンパク質機能予測に使われています。

生物学研究における深層学習の応用例としては、

- **タンパク質の構造予測:** AlphaFold2などのツールは、アミノ酸配列からタンパク質の立体構造を高精度に予測します。
- **遺伝子発現解析:** 遺伝子発現データから、疾患の診断や薬剤応答性の予測などを行います。

- **顕微鏡画像の解析:** 細胞や組織の画像を解析し、自動的に異常を検出したり、細胞の種類を分類したりします。

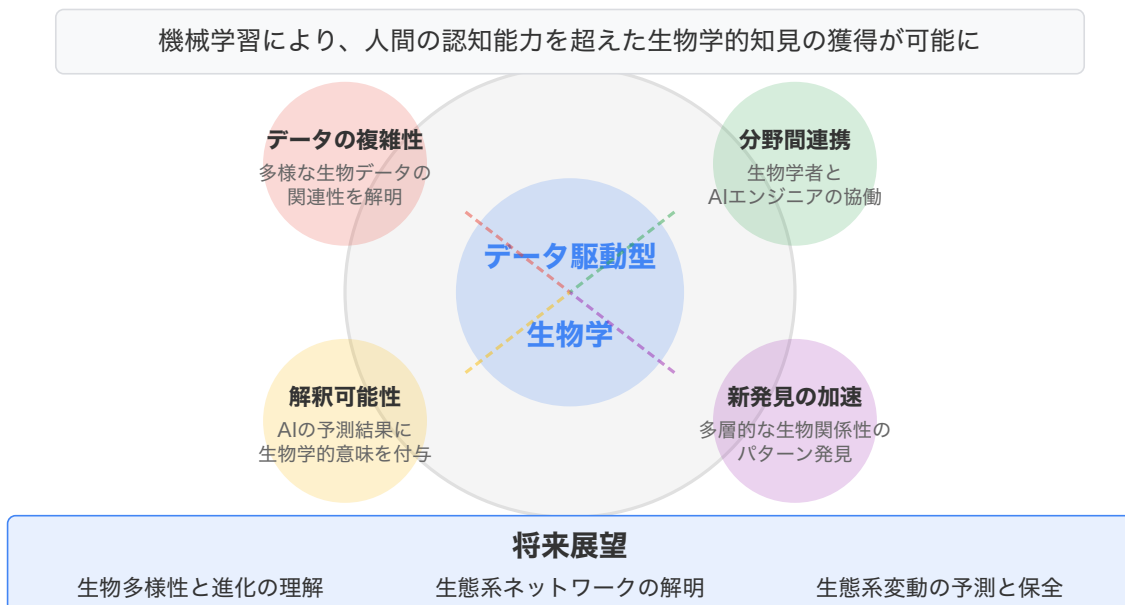
などがあります。

## まとめ

機械学習とは、コンピュータがデータから学ぶ技術です。生物学の研究では、生態系モデリング、ゲノム解析、タンパク質構造予測まで、様々な分野で活用されています。

機械学習を活用することで、これまで解明が困難だった生命の謎に迫る新しい発見がどんどん増えていくことが期待されています。

## 機械学習が拓く生物学の未来



ちなみに…

この配布資料は、Gemini 2.0 Pro (Google) に下調べをしてもらって、ChatGPT 4o (OpenAI) に文章を執筆してもらい、Claude 3.7 Sonnet に文章の校正と図の作成をしてもらいました。主要な生成AIの得意分野を活かした連携プレイです。

尾崎はほんの少し修正を行っただけで、合計1時間弱で完成しました。