

## 人工知能

- 機械学習とは（アーサー・サミュエル 1959）

「明確にプログラムとして記述しなくてもコンピュータに学習させる能力を与える研究分野」

- データサイエンスとは

「データから知識と洞察を抽出する分野」

- 認知的状態に着目した分類

- 強い人工知能

人間の知能そのものを模倣することで人間と同じレベルの認知的状態を持つ人工知能のこと

- 弱い人工知能

人間の行動を模倣することで人間の能力の一部を代替える人工知能のこと

- 対象となる分野に着目した分類

- 汎用人工知能

設計時には想定していない状況にも対応できる人工知能のこと

- 特化型人工知能

設計時に想定した状況にしか対応できない人工知能のこと

- 発達レベルによる分類

- レベル1：制御プログラム
- レベル2：ルールベース
- レベル3：機械学習
- レベル4：深層学習

- 人工知能の歴史

- 第一次人工知能ブーム

- 1950年代 - 1960年代
- 囲碁や将棋など、狭いルールで運用されるゲームが研究対象
- 主に論理、推論、探索を行うことで効率的に勝つ方法が研究された
- 現実の複雑な問題を解くのは難しいことがわかり、ブームは終焉を迎える

- 第二次人工知能ブーム

- 1980年代
- 機械にたくさんの知識を詰め込むことで人間と同等の知能を得ようとした
- 人間の知識を機械に理解できる方法が研究された
- **エキスパートシステム**が登場した

- 第三次人工知能ブーム - 2010年頃から

- 生命の柔軟性を考慮した計算方法である**ソフトコンピューティング**が注目を集める

- 機械学習の研究が地道に発展
- ビッグデータの蓄積とクラウド計算が発展

## 機械学習

- 機械学習とは
  - 人工知能の分類の一つ。効率的かつ効果的にコンピュータに学習させる理論体系。
  - 認識や行動を決めるためのパラメータを、データを基に最適化する。
- 特徴量とは
  - 対象を判別するための数値情報。
  - モデルに入力する直接的な情報。
  - 作りたい人工知能に応じて人間が適切に作成する必要がある（**特徴量設計**）
- 教師あり学習とは
  - 正解を含むデータに対して行う学習法
  - 正解はラベルと呼ばれる
  - ラベルが存在しない未知のデータに対する精度を上げることが目的
  - 教師あり学習の対象は予測する対象の性質によって**分類問題**と**回帰問題**の2つに分けられる。
  - 分類問題とは、予測する対象（正解）が犬、猫といったカテゴリとなっている場合で、画像データなどを入力としたときに対応するカテゴリを予測する問題。大小関係や順序関係はない。
  - 回帰問題とは、予測する対象（正解）が実数値となっている問題のこと。分類問題とは違い、間の値に意味づけを行うことが可能で、連続値を当てる。実数値なので、大小関係や順序関係が成立する。
- 教師なし学習とは
  - 正解を含まないデータに対して行う学習方法
  - データの特徴を捉えたり、ルールや法則を捉えることが目的
  - 教師なし学習の対象は、主に**クラスタリング**と**次元削減**
  - クラスタリングとは似たデータ同士をグループ分けすること
  - 次元削減とはデータの持っている情報を適切に削減すること
- 強化学習とは
  - 最適な行動をするように学習すること
  - 正解が与えられるわけではなく、報酬という、行動に対する価値が与えられ、なるべく多く報酬がもらえるような最適な行動を学習するのが目的。
  - 正解があるかどうかの問題ではなく、行動を最適化するという点で、教師あり学習や教師なし学習とは目的が根本的に違う。
- 機械学習を活用する際の注意点
  - 過去データの存在
    - 機械学習では基本的にすでに起きたことに関するデータを学習させるため、過去に扱ったことがない事象に対する予測を行うことは基本的にはできない。
  - データの質と量

- 機械学習ではモデルの精度をよくするためにはなるべく多くのデータが必要となる。一般的に精度は量にほぼ比例する。ただ、質も重要で、いくらデータ量が多くても、ノイズだらけで本質が全然見えないデータを使っても当然精度はよくなる。
- データの定量化
  - 画像、音声、テキストなど、データには様々な形式があるが、これらは全て数値情報にまづ変換しなければならない。数値情報に変換ができないデータしか存在しないならば、別の方法を考えるか、別のデータに変える必要がある。
- 予測結果に対する説明の必要性
  - 機械学習は基本的に精度を上げることが重視され、なぜそのような予測結果になったか、に対する問いに答えることは得意ではない。扱っている問題に関して、なぜその結果になったのか説明が必要な場合は、機械学習ではなく、統計学を用いたほうが良い場合がある。

## 深層学習

- 深層学習とは
  - ニューラルネットワークをモデルとした機械学習のこと
  - 対象となる問題にとって最適な特徴量を自動的に抽出することが可能なことが特徴
  - 注目されるきっかけは2010年のILSVRCという画像認識コンペで従来の機械学習アプローチを大きく引き離して優勝したこと
  - 2015年のコンペでは、人間の認識精度を超えた
- ニューラルネットワークとは
  - 人間の神経回路を模したモデル
  - 入力層、隠れ層、出力層からなり、**ノード**と**エッジ**で全体の構造が表現される
  - 各層の重みが**パラメータ**となり、機械学習同様にデータを基に最適化が行われる
  - 入力層にデータを入力して隠れ層に順次伝えていき、出力層で最終的な出力を得る処理を順伝播という
  - 活性化関数は重み付き和の次に施される処理だったが、一般的に非線形な関数を使う。非線形とは直線の形で表せない性質のこと。例えば**sigmoid関数**や**Rectified Linear Unit(ReLU)関数**などが挙げられる。この非線形な性質を持つためにニューラルネットワークの表現力が増し、高い精度を達成することが可能となった。
  - ニューラルネットワークにおいては**誤差逆伝播法**により学習を行う。これは、出力と正解データとの誤差が後ろのノードへ伝播するように計算を行い、各層の重みを調整する手法。まず、順伝播を行い、各層におけるそれぞれのノードの値を保存しておく。次に出力層において誤差を求め、一つ前の層においてとるべき値を逆算し、その層におけるそれぞれのノードの値との誤差を求める、という手順を繰り返し、全ての層における誤差を計算する。これらの計算結果をもとにして出力と正解データの誤差の和である**損失関数**を算出し、これが小さくなるように各層の重みを調整する。
  - ニューラルネットワークの最適化とは**損失関数**がなるべく小さくなるような重み（パラメータ）を見つけること。損失関数とは正解と出力結果との間の誤差の和のこと。
  - 損失関数が小さい谷底は**局所最適解**と呼ぶ。損失関数が一番小さい谷底を**大域的最適解**と呼ぶ。
  - 大域的最適解の探索方法を**勾配降下法**と呼ぶ。局所最適解を避けつつ、大域的最適解を探索する方法については現状ではたくさん提案されている。

- オートエンコーダ

- オートエンコーダとは、教師なし学習を行うためのニューラルネットワークのこと。次元削減ができ、入力したデータと同じデータを出力する。
- オートエンコーダでは隠れ層のノードの数を入力層よりも少なくしておく。隠れ層のノード数が入力層のノード数よりも少ないため、情報をそのまま出力層に伝えることができない。このような制約の中で、学習を行う際は出力層において入力層のデータが復元されるようにする。つまり、少ない情報量でデータをうまく復元できるように学習を行う。結果として隠れ層における出力が入力の次元削減された結果となる。
- 順伝搬の流れとしては、まず入力層から隠れ層に至る処理が、次元を圧縮している部分で、\*\*符号化(エンコード)\*\*と呼ぶ。隠れ層から出力層に至る処理が圧縮された情報からデータを復元する部分で、\*\*復号化(デコード)\*\*と呼ぶ。このように符号化と復号化を連ねることで自動で最適な符号化方法を学習する過程がオートエンコーダというアルゴリズム名の由来。

- オートエンコーダの種類

- DAE (Denoising Autoencoder)
  - 通常のオートエンコーダとは違いあえてノイズを加えたデータを入力として元のデータを復元するオートエンコーダ。
  - ノイズを除去して元のデータを復元しようとするため、通常のオートエンコーダよりもノイズに強い次元圧縮が可能となる。
  - 結果としてノイズを除去したデータを生成するので、データのノイズを除去したいときにも利用可能。
- VAE (Variational Autoencoder)
  - 隠れ層における出力の平均と分散を計算し、その結果を利用して新しい特徴量を生成し、入力とは違う新しいデータの生成を可能にするオートエンコーダ。
  - 新しいデータを生成することができるため、VAEで作成したデータを疑似的に教師あり学習の学習用データとして利用することも可能。

- 深層生成モデル

- 敵対的生成ネットワーク (GAN:Generative Adversarial Network)
  - 次元削減やクラスタリングを行うわけではなく、存在しないデータを新たに生成することが可能。生成器と識別器の、2つのニューラルネットワークからなる。 - 生成器はノイズや抽象データを入力して新しいデータを生成する。 - ノイズを入力する場合はノイズのパターンに応じて様々なデータを生成させ、抽象データを入力する場合はある程度狙ったデータを生成する。 - 識別器は生成器で生成されたデータと元のデータを入力して生成されたデータが本物か偽物かを識別する。 - 学習させる際は、生成器ではなるべく本物に近いデータが生成できて、識別器ではなるべく本物のデータかどうかを識別できるようにする。

- 深層強化学習

- Q学習
  - Q学習とはQ値と呼ばれる行動の価値を学習する方法
  - 目先の報酬や将来得られる報酬の期待値などをもとに定義する。
  - 対応付けは表の形でデータとして保持することになるが、状態や行動の数が増えると爆発的に大きくなってしまう。そこでこの対応付けにニューラルネットワークが利用される。

すなわち、状態を入力としてそれぞれの行動に対する価値が出力されるようなニューラルネットワークとなる。

- 一般にDQNと呼ばれる。元々Googleの子会社ディープマインドが開発したアルゴリズム。
- 対応表をニューラルネットワークに置き換えることで、より柔軟に様々な状態に対応したそれぞれの行動の価値を得ることができるようになる。
- ただし、とりうる行動の数が多くなると学習がうまくいかないことがある。
- Q学習は、Atari2600のゲームに適用し、49種類のゲームのうち、43種類で人間を上回った。しかし、深い先読みが要求されるゲームには弱かった。

○ 方策勾配法

- ある状態においてどの行動をどのくらいの確率で行うかを求める。
- 状態を入力として各行動の確率を出力するようなニューラルネットワークを作る。
- 行動を繰り返して、状態、行動、報酬に関するデータを収集し、これらをもとに学習データを作成する。
- 高い報酬の行動をとる確率が高くなるように、低い報酬の行動の確率が低くなるように学習を行う。
- 一般的に安定した出力結果を得るまでの時間が短い一方で多くの学習データが必要となる。
- 方策勾配法の適用例としては、ディープマインドが開発した以後プログラムAlphaGoがある。