

## 第 90 回ブレイクスルー研究会議事録

### 「生成 AI の研究の進展と社会へのインパクト」

日時 2023 年 6 月 28 日（火）18 : 00 - 20 : 10

会場 オンライン開催 ZOOM

参加者 20 名

講師 鈴木雅大特任助教（東京大学工学系大学院松尾研究室）

講演の内容

#### 1. 人工知能の研究の歴史

- 人工知能の定義は知能のとらえ方で人さまざまだが、最先端について使われることが多い。
- ダートマス会議で名前が登場してから、第 1 次、第 2 次、第 3 次と波があり、その間に冬の時代が存在してきた。「フレーム問題」「シンボルグラウンディング問題」を経験してきた。
- 学習により「汎化」を行えることが肝要である。
- 機械学習では、関数を求めることがターゲットであった。パラメータを変えてゆく。過学習を避けることが必要である。ここでは、教師あり、教師なし、強化学習に分類される。
- 深層学習では多層のニューラルネットを用いる。入力側からの順伝搬モデルが一般的である。情報の伝搬に重みを付ける。この重みを変更することで学習が行われる。
- 多層パーセプトロンが一時期重んじられたが、これも冬の時代を迎えた。
- オートエンコーダによる事前学習の導入などが検討され、現在の深層ニューラルネットワークにつながった。
- 深層ニューラルネットワークの内部では表現学習が行われ、入力を分解して組み上げるプロセスが確認できた。一方で、深層ニューラルネットワークをだますことができるような入力が存在することが判明した。

#### 2. 生成 AI の研究動向

- 生成 AI では三つのモデルが研究対象である。深層生成モデル、基盤モデル、世界モデルである。
- ① 深層生成モデルでは、2014 年頃から研究が始まり、画像が主に対象となっているが、文書や音声などにも使われている。例えばイヌの画像を集めて、確率モデル（生成モデル）を学習し、そのモデルから生成することで妥当なものとしてのイヌらしい画像を作ることができる。深層生成モデルでは生成モデルに深層ニューラルネットワークが使われている。
- 生成モデルと離れているものを見つけることで、異常検知や外れ値検出ができる。

また、それをモデル側に戻すことで、欠損値補完などもできる。

- 潜在変数 (Latent Variable) と呼ばれるデータの要因となるものを求める。推論と生成がニューラルネット上で行われる。
  - 多くのモデルがあるが、有名なのが VAE (Variational Auto-Encoder) がある。推論を encoder が、生成を decoder がおこなう。他に、GAN や拡散モデルがある。
  - VAE や GAN などでは、潜在空間での表現学習を行う。これを用いると画像の演算などが可能である。
  - 入力に対してラベルをつけないモデルが多いが、ラベルで条件付けることも可能で、これを条件付き生成モデルと呼ぶ。これを用いると文章から画像を生成できる。
  - マルチモーダル学習では様々な種類のデータ (人間でいうと五感から得られる情報) を通じて学習を行う。深層生成モデルを用いたマルチモーダル学習では、異なるモダリティ (例えば画像と属性) 間を双方向に変換することができる。
- ② 基盤モデルは、GPT などがそれにあたる。自然言語を中心に Transformer という技術が用いられる。Transformer には自身の単語間の類似度を学習する self-attention という構造が含まれている。
- 大規模言語モデルでは、自分自身を教師とする自己教師あり学習という方法で学習する。
  - いろいろ使ってみると、予想よりも何でもできることが分かった。
  - 同じモデルで、パラメータを更新しなくても、プロンプトを変えるといろいろなタスクに対応できてしまう。
  - 数学や論理の問題を解くことは苦手である。
  - 順番に考えよとプロンプトを与える (step by step) と数学や論理の問題が解ける。Chain of thought と呼ばれる。
  - 単一のモデルでいろいろな問題が解けるので基盤モデルと呼ぶ。言語も画像も使える。
- ③ 世界モデル
- 松尾研究室で力を入れているモデルである。
  - 我々はどのようにして世界を知覚しているかを扱う。例えば、人は写真を見ても 3 次元であることを認識できる。これは、我々が写真という結果から、経験に基づきその原因を常に推論しているからであり、ヘルムホルツの無意識的推論といわれる。
  - この経験をモデルとして脳内で学習によって獲得している。このモデルを世界モデルという。世界モデルの研究は、これを人工知能で実現することを目的とする。
  - 世界モデルを実現するためには、外界からの情報を空間的・時間的な表現に圧縮する必要があるため表現学習が重要である。そのため深層ニューラルネットワークを用いる。
  - 世界モデルで推論の他にもう一つの重要な概念が予測である。これは、推論とは逆

の流れで、ある原因のときに、どのような結果となるかを予測する。野球で、打者は投球を見てバットを振ったのでは間に合わない。予測があるのではないか。錯視も似た現象である。

- 頭の中でもシミュレーションを行っている。これはロボットに応用できるのではないか。

### 3. これからのAI

- システムには、システム1とシステム2があるという。前者は直感的、後者は時間がかかるが論理的な計画などにあたる。
- AIでは、認識、行動、言語の順に技術が進歩していくとみていたが、基盤モデルで一気に言語に適用できてしまった。
- 産業への適用が期待できる。

### Q&A

多くの質問やコメントが出された。

確率が多く用いられる。計算機の機能限界が制約になるかもしれない。日本ではTransformerを研究していたが計算機費用の限界で多数展開ができず実用化できなかった。GPTなどの効果をわかりやすく示したい。GPTでは潜在変数を明示的に定義していないが自然に求まっているのかもしれない。あるタスクが他にも使えるのがおもしろい。多くの仕事はAIに代わられるが目的を設定できるのは人間である。

以上