

近未来金融システム創造プログラム第 10 回講義レポート

第 10 回目となる本日は、東京大学大学院工学系研究科の松尾豊教授から「社会と産業を変える人工知能の未来」という題目で講義が行われた。ディープラーニングの技術的な内容だけでなく、ディープラーニングが活用される分野や、今後の発展や社会への影響などについても語っていただいた。

ディープラーニングの進展

近年、第 3 次 AI ブームの中で、ディープラーニングの研究が盛り上がっている。ディープラーニングとは、多層のニューラルネットワークを用いた学習手法の総称であり、それぞれのユニット（ニューロン）が、重み和（重みをかけた合計を求める処理）と活性化関数（0-1 への丸め）の処理を行う。また、入力値に対して目的とする値が出力されるよう、出力値の誤差から重みを調整する誤差逆伝播を行う。このようにして、AI の精度はディープラーニングにより向上していく。ディープラーニングが最も端的に使われる分野は画像認識であり、製造における不良品検知や自動運転、インフラの点検、医療画像などのさまざまな産業で、ディープラーニングが活用されている。

今までは AI が活用される分野は画像認識が主流だったが、ディープラーニングの登場によって自然言語処理の性能も飛躍的に向上した。そのベースとなったのが、トランスフォーマーと呼ばれる技術である。2017 年、「アテンション」という機構を大規模に使ったトランスフォーマーが考案された。トランスフォーマーは従来のニューラルネットワークとは異なり、重みだけでなく、注意の当て方自体も学習できるようになり、柔軟で精度の高い処理が可能となった。また、自己教師あり学習を行うことによって、少ないサンプルからでも精度が大きく向上するようになった。2020 年に登場した大規模言語モデルの GPT-3 は 1750 億パラメータを持ち、それまでは困難と言われていた、極めて自然な質問応答や文章生成を行うことができた。パラメータは、数が増えるほどにエラーの出る数が減り、精度が増すという法則（スケール則）がある。そのため、パラメータのより多いモデルの開発競争が起こっている。

近年、画像生成の分野においても、ノイズのかかった画像から綺麗な絵を生成する操作を AI に学ばせることで、文章で入力された情報をもとに画像生成させることが可能になった。最近では、画像生成の面で AI の技術が評価されており、自然言語認識の技術はあまり評価されていない。しかし、画像認識も自然言語処理も、いずれもトランスフォーマを使ったモデルが基礎になっており、ファウンデーションモデルと呼ばれる。日本はまだその重要性を

認識していない。

ディープラーニングは今後ますます進歩していくと思われる。なぜなら、今の深層学習にはいくつかの根本的な問題があり、そのせいで達成されていないタスクがあるからだ。今後予想される進歩を考える上で、ヒントになるものを紹介する。1つは、リザーバーコンピューティングである。リザーバーコンピューティングのリザーバー層では、ネットワークがランダムに配置されており、たまたま意味を持った組成が出力の際にかけられる重みは大きくなるため、AIの精度は高くなる。

2つ目に、ディープラーニングでは、データが多く必要とされるが、パラメータより少ないデータから学習している。これは、本来おかしいことであるが、その理由は解明されていない。それに対して、2019年に宝くじ仮説という学説が発表された。ある程度の精度を持った学習済みのネットワークで、かけられている重みが軽いコネクシオンの多くを刈り取って、初期値に戻して再学習させても精度が下がらないことが知られていた。しかし、刈り取ったコネクシオンの初期値を変えて初期化すると精度は上がらない。「初期値とネットワーク構造の組み合わせ」が宝くじであり、そのなかでたまたま当たりくじを引くと精度が高くなるということではないかという仮説が宝くじ仮説である。つまり、宝くじはたくさん買った方が当たりが出やすいのと同様に、パラメータも多い方が当たりが出やすいため、パラメータはデータの数より多くされるのではないかということである。さらに、強い宝くじ仮説では、学習すら不要であると主張されている。

3つ目に、モデルを大きくしていくと、初めはエラーが減っていくが、後に上昇に転じることがわかっていた。しかし、そこからさらにパラメータの数を増やしていくと、これまでに知られていた領域の先に、Overparameterizedでエラーの少ない領域があることがわかった。これがDouble Descent (2019)である。これは、「与えられた構造の中で重みを調整する」モードと「良い構造を探索する」モードがあり、これらのモードが切り替わるため、エラーの減少が2回起こると解釈できる。また、Grokking (2022)では、訓練の精度が上がった後、時間がたってから急にTestの精度があがることが示された。これは、おそらくコンピュータが「良い構造」を見つけ出すことでTestの精度が上がっており、「与えられた構造の中で重みを調整する」ことより「良い構造を探索する」ことの方が時間のかかる難しい学習であることがわかる。

このように、決められたニューラルネットワークの構造における重みの調整(学習)というのは、従来から行われている機械学習であるが、それ以外に「構造の探索」という、より重要な学習があるらしいことがわかる。これは、深層学習が従来の機械学習とは決定的に異なる点であるかもしれない。

パラメータの多いモデルほど精度の高い学習ができるが、今までの科学では、パラメータの少ないモデルが美しいとされ、パラメータを増やすことは避けられてきた。人間の認知に限界があったために、少数パラメータのモデルが好まれてきたと考えられる。しかし、ディ

ープラーニングは膨大なパラメータを持っていても上手く動き、近年では人間の理解の限界を超えていたとしても、より予測精度の高いモデルを許容する風潮が生まれてきた。

「理解」とは何か

では、人間が「理解する」とは何か。人間の脳は多数パラメータを扱い、モデル化を行うことができるが、「理解」した気にはなっていない。「理解する」ということが何を意味するのかを考えるためには、人間の知能の仕組みを知る必要がある。脳は世界モデルを学習した後、言語の入力を受けると、入力内容が画像生成 AI のように絵で描き出され、言語による伝達が可能になる。同様に、脳が記号を学習することで記号による伝達も可能になった。理解とは、言語を通して人間が扱えるもののみで起こり、知覚を通して知っているが言語化できないものは「理解している」とは言わない。同じ初期状態を設定し、同じ操作で、同じ心的な状態を再現することが「理解」の肝である。

知能に関して、人間は自分たちが最高の知能を持っていると思い込んでいるが、人間の知能は、他の動物などの数ある知能の 1 つにすぎない。ディープラーニングの進歩により、人間にとって自身の知能の相対化が可能になることを期待する。

Q&A

Q1. 医療システムの AI や自動運転の AI が事故を起こした際、AI は責任をどのように取るのか。

人間が責任を持つ際は、問題を起こすと自身も損失を被る人物が責任を持つことで、信頼が発生する。しかし、AI はこのような形で責任を持つことはできない。そのため、良い仕事をしてきた AI に信頼を置き、より重要な仕事を担わせていくことで、擬似的に責任を持たせることができる。

Q2. AI に対して、法律面ではどのような対応がとられるか。

法律は、AI の進展に合わせて大きく変わってくるだろう。日本は社会的規範意識が強く、グレーとされる技術の発展は遅い。一方、アメリカや中国では、グレーであってもテクノロジーの開発は大きく進められるため、テクノロジーが進んだ上で、後から生じた問題に合わせて法律を整備するという形がとられている。日本はテクノロジーの進展で後れを取っているため、アメリカや中国同様、あまり固く規制せずに開発を進めていくのが良いと思われる。

Q3. 人間を超える人工知能は、個々の分野で人間の能力を超える人工知能を組み合わせることで作れるか。

優れた人工知能を組み合わせた統合的な知能は登場すると思われる。

Q4. 人間の能力を超えた人工知能を社会で活用するためには、人間の理解できるレベルにまで落とし込む必要があるため、人間が理解するための説明が必要になるのではないか。

説明という行為の本質は、説明する努力をすること自体が評価され、説明する内容の正当性は重要視されない。これは、社会的な生物としての人間に備わった性質によると思われる。そのため、「説明」という行為自体が必要でなくなる可能性がある。

Q5. 芸術の領域でAIはどこまで優れた作品を作れるのか。

人間が、モノをどのように感じるかは進化の過程で規定されてきた。しかし、文明の発達によって人間はこの規定を乗り越えることができた。一方で、そのような効用の大小を見極める機能を持たないAIが、優れた作品を作ることは難しい。しかし、作品を作り人間の反応から学習して作品を最適化することはできるため、AIが人間の想像の及ばない作品を作り出し、それが面白い作品だとされる可能性もある。

Q6. 天気予報はあまり当たらないが、ディープラーニングの活用により精度が上がると思われる。しかし、いまだ天気予報の精度に改善が見られないのはなぜか。

ディープラーニングにより天気予報の精度が上がることは実証されており、試験的な運用は既に始まっている。しかし、天気予報の精度改善には金銭的なリターンがあまりないため、実際の導入はいまだ進んでいない。