

近未来金融システム創造プログラム第9回講義レポート

第9回目となる本日は、東京大学名誉教授で理化学研究所脳科学総合研究センター名誉研究員の甘利俊一先生から「機械学習の発明」という題目で講義が行われた。人工知能がどこから生まれ、どのようにして今の形になってきたのか。人工知能の歴史と、それが社会に及ぼす影響、そしてわれわれ人間とどう違うのかについて、お話しいただいた。

生命の仕組み

「人工知能は人間の脳を超えるのか?」という問いを議論するためには、人間の脳と人工知能の類似点・相違点について考えることが重要だ。

では脳はどのような仕組みになっていて、どうやってできたのか。

人間の脳を理解するためには、138億年前のビッグバンまで遡る必要がある。ビッグバンによって、突如として高密度の高エネルギーが生まれ、それが物理法則に従って発展してきた結果、地球が生まれ、生命が生まれた。生命は物質でできているが、その中に物質を形作る情報も書き込まれている。そしてその情報を使って、自分と同じものを複製することができるが、複製を間違えることもある。その際、より強靱な生命体が誕生することがあり、これが生命の進化の法則である。

我々人類はこの流れの中で誕生し、外界の情報を受容・処理し自分の行動を指令する部分として、脳神経系を発展させてきた。脳は非常にたくさんの神経細胞が繋がったシステムとして機能している。それによって我々は、心や意識、知能を備えるに至り、合理的な行動をとるようになった。その帰結として作られたのが、社会・文明である。

人工知能の歴史

コンピューターの起源は、1936年にイギリスの数学者であるアランチューリングが考案したチューリングマシンである。チューリングマシンは与えられた問題について、解を求めるアルゴリズムがない問題には解答できないものであった。その後、コンピューターは戦争に利用するために開発が進められ、第二次世界大戦後実用に供されるようになってきた。

第1次人工知能ブームと言われる1950年代には、コンピューターに人間と同じような知的な作業を行わせることができるのではないか、と考えられ始め、プログラムを与えてコンピューターに高速で実行させ、知能に似たような仕掛けを実現しようという動きが生まれた。同じ頃、心理学者のローゼンブラットによって Perceptron というモデルが考案された。これは、人間が学習によって徐々に知能を獲得するように、コンピューター上で脳のようなモデルを作って学習させれば、人間が解答可能な問題を全て答えられるシステムが作れるのではないか、というものである。しかし、どちらの試みも当時の技術では実用には程遠く、廃れてしまうこととなった。

それから約20年後の1970年代には、トランジスタの誕生でコンピューターの性能が格段

に上がっていたこともあり、第2次人工知能ブームが訪れた。この時期、専門家の知識を大量に蓄積し、コンピューター上に書き込まれた推論の規則に従って問題を解いていけば、知的な機能が生まれるのではないかと、というエキスパートシステムが誕生した。実際に MYCIN という感染症を診断するプログラムが考案され、町医者程度の診断精度は再現できたものの、装置が大きすぎるなどの理由から実用には至らなかった。一方で、神経回路網を使ったモデルについても格段の進歩があったが、こちらもコンピューターの性能の限界に突き当たってしまい、実用に供するような知能は生み出せなかった。

2000年代に入ると、コンピューターの処理機能の多層化により、データ処理速度が向上していたことなどを背景に、コンピューターの性能が格段に上がり、情報を蓄えるデータベースも改善された。こうした状況の中で、ヒントンによって生み出されたのが深層学習という考え方である。これが火付け役となり、2010年代から現在まで続く第三次人工知能ブームが巻き起こった。人工知能は初め、囲碁や画像認識などの静的なパターン認識で人間を凌駕していったが、次いで動的な言語の学習・生成などについても、同じようにパターン認識を行えるようになり、現在は ChatGPT をはじめとする生成 AI が誕生している。

深層学習とは

脳においてニューロン(神経細胞)は、ほかの 1000~10000 ものニューロンから入力を受け取り、その情報を総合して計算して答えを出し、他のニューロンに伝えるという、非常に複雑な仕組みを持っている。深層学習においても、Perceptron という人工的なニューロンのモデルを用いて、ニューロンのつながりを数理モデル化した層状学習回路網に情報を入力し、それぞれの神経細胞で計算を繰り返して答えを出す仕組みとなっている。

一般的な深層学習には、確率勾配降下学習法が採用されている。ここでの学習は、正解の分かっている問題を層状学習回路網に入力し、正解に近づくようにニューロンに対する入力のパラメーター(重み)を調整していくことで行われる。

神経回路網で計算機械を学習で作る際には、様々な問題が生じる。1つは局所解の問題であり、局所解を得られたときにパラメータの修正が止まってしまう。ところが近年になって、神経回路網の学習可能なパラメーターの数を大きくすることによって、この問題を解決できることがわかってきた。

他にも汎化誤差の問題がある。パラメーターの数がある程度大きくなると、不必要なノイズが増えてしまう過学習の問題が発生し、汎化能力が下がってしまう。しかし、その域を超えてさらにパラメータを増やしていくと、汎化誤差が少なくなることが近年明らかになってきた。

深層学習についての理論は大変遅れており、局所解や汎化誤差の問題もまだまだ完全な解決を見ていない。ただ少しずつは進んできており、最近ではニューラルタンジェントカーネル関連などの理論も話題になっている。

人工知能と人間

Libet の実験により、人間は脳が計算を行い、判断して指令を出してから、その判断内容を認知していることが明らかになった。つまり脳が判断を行い、意識はその判断が他の様々な情報と矛盾しない合理的なものかどうかを考えるために利用される。人工知能にはこのような機能はない。

人間は心を持っていて、心で決断をくだす。当然、喜びすぎたり、悲しみすぎたりすることによって、合理的な判断を行えないこともある。一方、ロボットが悲しんで仕事をしなくなったら困るため、そういう意味でロボットが心を持つことはない。人間から見たときに、ロボットが心を持っているように見せかけることはできるかもしれないが、その際ロボットは、全て計算づくで自分の心を決めることになる。つまり人間のために喜ぶふりをしたり、悲しむふりをすることはできるが、それは全て、その方が良いという計算から導き出された結果である、ということである。

現在、人工知能は安全なのか、もっと制御していくべきなのか、議論がなされている。ヨーロッパを中心に、人工知能の暴走を制御しようという議論があるが、暴走するのは人工知能ではなく、人間である。人工知能を使って軍事に活用したり、よくない金儲けに悪用したりするのは人間であるため、いかに利用者に規制をかけるかが重要になってくるだろう。人工知能の性能が将来格段に上がっていくと、人類の生産性も格段に上がるが、同時に人々の格差をより拡大してしまうかもしれない。対策としてベーシックインカムを導入する案があるが、それに甘えてただ寝て暮らしているのでは、人間の家畜化が進んでしまう。必要なことは、人間に働く喜びを与えることである。各人が自由で、平等で、協調する社会を作っていくためには、人工知能の優れた生産力を使って何をすべきか、ということを実際に考えていかなければならない。

Q&A

Q1

人間が思考をする際、エネルギー消費量は非常に少ない。一方で大規模言語モデルは、人間らしい回答を出せるようになっているが、CPU の計算にエネルギーがとてかかる。その違いは何なのか？

A1

エネルギー効率だと、人間の脳はだいたい 25W である。一方で、コンピューターについて考えると、例えば理研の「富岳」は小都市の全世帯の電力に相当するような電力を使っている。要するに、コンピューターは電力がいくらかかっても、何もかもを計算するが、人間は脳をそこまで使わなくても良いため、エネルギー効率は桁違いである。その仕掛けが分かっ

たらずごい。

電力の問題は今の技術では解消できていない。ニューロモルフィックシステムのような、ニューロンに似た動作をするシステムが完成しないと、電力問題は解決しないのかもしれない。

Q2

ChatGPT の性能に驚かれています。長年研究されてきた先生にとって、大規模言語モデルは、どのあたりが驚愕したポイントだったのか？

A2

試しに、「素数とは何か」という質問を ChatGPT にしてみると、しっかりと答えが返ってきました。「素数は何個あるのか」という質問をすると、それにもしっかりと答えが返ってくる。ただその証明は正しくないものであったが、ChatGPT にはそのような豊富な知識がある。素人では一見わからない、真実でない部分が含まれている可能性があることは怖いですが、ChatGPT 自体は原理を理解していなくとも、それにまつわる性質や過去の歴史等の知識が非常に豊富である。

Q3

ロボットは心を持たないという話があったが、データをたくさん学習させることによって、ChatGPT もある程度心を理解できるようになるのではないかと思う。ChatGPT が人の心を数値化したり、カウンセラーの役割を果たしたりなど、人間以上に心を上手く扱えるようになる未来は訪れると思うか？

A3

可能だと思うが、それはロボットが心を持つということではない。様々な医療や心に関する知識をいっぱい持っていて、こういう場合にはこういった対症療法が必要という判断ができるようになれば、変な人間よりもよっぽど役に立つ様になるかもしれない。それが良いことなのか、恐ろしいことなのかについては、使う人間次第だと思う。

ChatGPT を鵜呑みにしたり、それに頼るような社会や文明を作っていくと、落とし穴にはまるだろう。コンピューターの間違いは今後どんどん少なくなっていくだろうが、その後で ChatGPT が総合的な人格を持てるのかどうかは疑問である。人間は進化の歴史で人格を作ってきて上手くいっているが、もちろん完全なわけではない。例えば ChatGPT が人格を持ったときに、それが悪人の人格だとしたら非常に困る。人工知能が知識を持っていて、人間がそれを利用することは非常に良いが、コンピューター文明を人間がどう制御していくのか、これは大きな問題だと思う。

Q3

・機械学習で情報を解析している中で、インプットやアウトプットが増えてきている。書式

やフォーマットが限定されている中で、新しいカテゴリーが出てきても対応できないのではないか、という懸念があるが、それについてはどうお考えか？

・一昔前までは、AI がチェスや将棋の名人に勝った、という話が話題の中心であったが、近年、急に ChatGPT などが驚愕するようなアウトプットを出すようになってきた。素人からすると、急にブレイクスルーを果たしたような印象を受けるが、そのきっかけとなった出来事や技術革新、背景は何なのか？

A3

新しいカテゴリーについては、初めから新しいカテゴリーが出た時にどう対処するのかという仕組みを仕込んでおけば、それなりに解決できると思う。AI はどんどん進化していくだろうし、今の単純な学習システムが万能ではない。

ブレイクスルーについては、深層学習が一番の基礎にある。それに強化学習や、Diffusion Model、トランスフォーマーモデル、Attention などが絡んでいる。これは自然な流れではあるが、計算力・電力・人力について、ものすごい物量作戦で為されてきた背景がある。非常に驚いているのは、自分たちすら気付かぬうちに、今まで想像もしなかったようなことが、可能になっているということである。つまりシステムが新たな仕組みを創発したのかもしれない。人間は後追いで、そのような仕組みがなぜ起こるのか解明しようとしているが、その解明は非常に遅れている。

Q4

学者の視点から見ると、人間が AI の暴走を防ぐことは本当に可能なのか。また、AI の暴走を防ぐために、どういう対策が必要なのか？

A4

AI の暴走を防ぐということは、AI を規制するのではなく、人間を制御することに尽きると思う。現在、生成 AI に関して様々な規制をかけようとしているが、それで生成 AI の活動が止まることはない。そうではなく、人間をどう制御するか、問題が一番あるのだと思う。