

「共感」する人工知能

京都大学文学部第7講義室

2015年4月17日(日)13:00~14:30

0 はじめに

第三次人工知能ブームと呼ばれる現代において、人工知能研究の目標の一つは仮説形成（アブダクション、abduction）であるとされる。例えば、高橋恒一先生（理化学研究所 生命システム研究センター）にお話を伺った際には、差し当たり目的は機械による仮説形成である、と公言されていた。少なくない研究者がコンピュータによる仮説形成は可能になり、いずれ「研究そのもの」もコンピュータによってなされるだろうと予測している。研究者はクリエイティブな職業だと油断していると、足をすくわれることになりかねない。

しかしながら、「2016年はコンピュータが感情を把握し始める年になる」とも言われ、コンピュータによる感情把握が現在、特に盛り上がっている話題である。日本でも、例えば栗原聡先生（電気通信大学大学院 情報システム学研究科教授）は、感情こそが今後の人工知能研究において重要であるとして、場の空気を読む人工知能システム・Aung-AI（阿吽・人工知能）を開発されている。今回の講義では、学問の鍋の専攻多様性の文化を尊重しながら感情と人工知能をテーマに考えていきたい。そして最後に、感情ではなく「共感」についてこそ考えなければならないという、講演者の主張を紹介させていただきたい。

1 人工知能とは何か？

まず、議論をするために必要なことは、言葉の定義、前提条件の共有である。今回の議論を始める前に、人工知能とは何なのか考えたい。「人為的に作られた、インプットに対して何か別のアウトプットを出すもの」（参加者より）

その他の定義も見て、人工知能の定義を決めよう。Wikipedia 日本語版によると、「人工知能（じんこうちのう、英: artificial intelligence、AI）とは、人工的にコンピュータ上などで人間と同様の知能を実現させようという試み、或いはそのための一連の基礎技術を指す。」

ここで「人工」を「人為的な」と読み替えてやると、知能 = 試み、または知能 = 技術となり、論理構造に問題がある。後者はまだしも前者は疑わしい。

ここで、人工知能学会「What 's AI」を参照しても、立場の議論にすり替えることで明確な定義を避けていることがわかる。

『人工知能』とは何だと思うのでしょうか？まるで人間のようにふるまう機械を想像するのではないのでしょうか？これは正しいとも、間違っているともいえます。なぜなら、人工知能の研究には二つの立場があるからです。一つは、人間の知能そのものをもつ機械を作ろうとする立場、もう一つは、人間が知能を使ってすることを機械にさせようとする立場です。

人工知能の二つの立場が「人の知能をもつ機械を作る」「人が知能を使ってすることを機械にさせようとする」ということを紹介しているが、相変わらず定義はできていない。

実は、特に日本語の文献だけに問題があるのではない。「人工知能」に特定の定義がないことは問題として知

られている。浅田稔先生（大阪大学大学院 工学研究科教授）も人工知能の定義は困難であると述べている。そもそも人工という言葉の定義は分かって、知能の定義が曖昧であり、それゆえ人工知能の定義が難しくなっているためであるという。

ここでは、国語辞典にならって、知能を「人間のように理解し、思考する能力」と定義付けることによって、人工知能を「人為的に作られた、人間のように理解し、思考する能力」と定義することにし、仮に議論を進めてみよう。

2 人工知能の種類

人工知能について定義したが、今度はその種類について考えたい。人工知能の種類の分け方として、「強い」と「弱い」に注目した分け方がある。その分け方の一つを紹介する。

- ・強い人工知能:汎用性をもつ
- ・弱い人工知能:人工知能の部分的再現
(超知能:生物的、集合的超知能など)

強い人工知能が汎用性をもつということであれば、現状の全ての人工知能が弱い人工知能に分類される。ただし、超知能を考慮すると、身近な電卓などでさえ超知能を持つとも言える。仮想世界の中の話だけでは、人の知能とはほど遠いのではないかと指摘する研究者も多い。端的に言えば、AlphaGo は汎用型に、近づいたとはいえないという主張である。そう考えると全てがほとんど全てのものが弱い人工知能であり、また、強い人工知能と超知能の違いも曖昧である。

そこで、もう一つの人工知能の分類を紹介しようと思う。

1. 強い人工知能：感情を理解し、意識をもつ
2. 弱い人工知能：感情を理解せず、意識を持たない

冒頭でも述べたように、感情理解は人工知能研究の最前線でもある。CMU（カーネギーメロン大学）の計算機科学部長の Andrew Moore は、「2016 年はコンピュータが感情を把握し始める年になる」と予測している。

上述の「強い人工知能」実現に向けての研究は、次の二つの派閥に大別できる。機能実現派と構造再現派である。前者は数理的手法を用いるのが特徴で、Google DeepMind などその例である。後者は神経科学的手法を用いるのが特徴であり、欧州の Human Brain Project などその例である。これらは計算機性能の向上、新しい技術（深層学習など）などを取り込みながら、強い人工知能を実現しようと試みている。

また、よく語られる議論の一つが、Alan Turing と Marvin Minsky の対立である。これについては詳しい内容は参考文献に譲るが、大事なことは、大局的に見ること、すなわち、極度に数学的でもなく、極度に神経科学的でもなく、中道を行くべきということかもしれない。

3 身体性の議論

脳を模すことで知能は生まれるか？

1. 数理的な機能実現は成功するが、擬似的な知能であることには代わりがなく、ある種、人を騙しているということは変わらない。
2. 対して、構造再現は失敗する。

「人の脳を取り出す思考実験」を検討すると、知能獲得の段階における身体性の必要性がわかる。

生きている人間の脳を取り出し、十分な栄養を与えることができている（生きている）と仮定すると、その脳は思考を続けているに違いなく、それはまた、人の知能そのものである。

すると、構造再現によって人の脳を再現することで人の知能が作れるということになるか？

少し考えれば簡単なことで、それは不可能である。赤子の脳を再現して放置したところでそれは成人の脳、すなわち、一般に言う知能には成長しないであろう。学習する段階では、人の身体を持って現実世界に触れてなければ人の知能は獲得できないということである。

これは松尾豊先生の言う子どもの知能の議論にも通じるが、上述のアイデアは基本的にはロドニー・アレン・ブルックスの「身体性」の議論であり、詳しくは「知能の原理」などの文献をご覧ください。

4 労働、身体性、人工知能の未来

人工知能によって仕事が奪われるという話はよく聞くことだが、人工知能やオートメーションによって人々の生活が楽になっているというのもまた事実である。便利や快適になればなるほど、人々はそれ以上のものを追いかけて、技術はそれに応えようと進化し続けている。だが、本当にそれらの技術は我々に恩恵をもたらすだけなのだろうか？「事実上労働の知る最も甘い夢である」と今から1世紀ほど前を生きていたロバート・フロストは言った。彼の主張の中で垣間見える、労働による実世界とのインタラクションにも注目したい。

森のそばで聞こえる音はただひとつ、わたしの長い大鎌が地面にささやく声だけだった。何をささやいていたのだろう。わたしもよくわからない。太陽の照りつける暑さについてだったか、何か、たぶん、音のないことについてか—だから、ささやくばかりで話さなかったのだ。怠惰は時間の恵みの夢でもなければ、妖精やエルフの手からたやすく得られる黄金でもなかった。真実より以上のものは弱すぎるようにおもわれたろう、ほっそりと先のとがった花（薄い色の野性蘭）も咲くなか、鮮やかな緑色の蛇をおびえさせ、湿地の草をなぎ倒す熱烈な愛にとっては。事実上労働の知る最も甘い夢である。私の長い大鎌はささやき、干すべき草を残した。

彼の主張を読み取ってみると、「我々は労働によって世界への理解を深め、事実を獲得してきた」と言えるであろう。機械にできることは機械に任せられるようになった。しかし、その結果として我々は人にしかできない知の労働によって事実を獲得しているであろうか？機械が進歩するほど人は思考を怠けるようになり、ただの機械の添え物に成り果ててはいないだろうか？技術は我々に何をもちたらし何を奪っていったのか、このことを今一度考えてみてはどうだろうか。

こちらでもまた、興味のある方は「オートメーション・バカ」などの文献を参照いただきたい。

5 「共感」する人工知能

最後に、「『共感』は感情のシンボルグラウンディングに先立つ」という自説を紹介する。

1. 人間はすべての感情をグラウンディングしていないが共感できている

ここでは、ほとんどの感情には名前がついていないが気持ちが伝わるというような意味である。例えば小説を読むとき、その複雑な感情には大抵の場合、名前も付いていなければ、ましてや、今までに同じ感情を抱いたことがないこともある。しかし、我々は共感しているということを経験している。

2. むしろ、共感することによって感情という記号が獲得される

これは浅田稔先生にご指摘いただいたことだが、どこまで生得的であるかは議論が分かれるものの、ミラーニューロンシステム(MNS)と同じアイデアで感情は獲得されると考える。共感によってしか感情の記号化は不可能であるという議論であり、一般的に考えればごく自然なアイデアでもある。

感情を獲得してこそ共感することができると言われると、まさにその通りと受け止めかねない。しかしながら、よく考えれば、(そもそも、感情の記号化ができるかどうかは保証されていないが、)共感することこそ感情獲得のための第一歩なのかもしれない。

そして何よりも、求められている人工知能は感情を持つ知能ではなく、共感できる知能なのだと考えている。

- 人工知能を脅威として捉える場合の危機は、人間に「共感」できる知能であれば回避できるためである

ここで思うのは、感情を持つ人工知能は必ずしもそれを回避できず、危険を増長する可能性すらあるということだ。ターミネーターなどの映画に代表されるシナリオである。

- また、「共感」できる人工知能となれば人は話したいと感じるはずである

共感のない対話は虚しいとすら思うし、人を導く人格者としての、共感という機能を備えた人工知能は、人を救うのではないかとすら考える。

6 まとめ

学問の鍋という様々な専攻の学生が集まる場にふさわしい、専攻の多様性に満ちた内容を心がけた。共感が感情に先立つならばシンボルグラウンディング問題は関係ない。また、求められている人工知能は感情の知能ではなく、共感できる知能である。講演者はこの「共感する人工知能」を心に秘めて、研究活動に取り組んでいるが、参加者はどう考えただろうか。冒頭で述べた通り、誰にとっても他人事ではないのである。そして、様々な立場から人工知能に着いて議論をすることができる。今回の講義を受けて、人工知能についての思うことがあったなら冥利に尽きる。