

人狼知能と人

東京大学大学院工学系研究科 准教授
鳥海不二夫 (とりうみ ふじお)



Profile—鳥海不二夫

2004年、東京工業大学大学院理工学研究科機械制御システム工学専攻博士課程修了。博士（工学）。2012年より現職。専門は計算社会科学・情報工学。著書は『人狼知能で学ぶAIプログラミング』（共著、マイナビ出版）など。

ゲームにおける

人間 vs コンピュータの歴史

いつから人間はコンピュータ（あるいは人工知能）と戦いたいと思いはじめたのだろうか。いや、この表現は正確ではないだろう。コンピュータはそもそも人間と戦いたいとは思わないものであり、人間がコンピュータに戦いを挑んだとき、その戦いに応えるコンピュータもまた人間が作らなければいけないのだから。その意味では、この問いは正しく言えば、「いつから人間はコンピュータに打ち負かされたかったのだろうか」となるのだろうか。あるいは、「いつから人間はコンピュータを使って他の人を打ち負かしてやりたかったのだろうか」となるのかもしれない。

世界初のコンピュータENIACが完成したのが1946年であるが、チェスのプログラミングに関する論文は1950年にすでに出版されている（Shannon, 1950）。コンピュータが開発された直後から人間とコンピュータの間での知恵比べは始まったと言えよう。また、コンピュータの父とも言えるラン・チャーリングは紙上でアルゴリズムを作成し、アルゴリズムに従ってチェスをプレイしたという（Turing, 1953）。当時はまだアルゴリズムを実装するだけの性能を備えたコンピュータは存在してい

なかった。

その後、演算能力の発展とともにコンピュータのチェスをプレイする能力は向上していった。人工知能批判で有名なヒューバート・ドレイファスは、「チェスには直感が必要だからコンピュータにはチェスは打てない」と豪語した上で、1967年にコンピュータチェスと対戦をして敗れた。つまり、この時点でコンピュータチェスはドレイファスよりは強かったと言える。そして、当時のチェス世界チャンピオンであるカスパロフにIBMが開発したチェスコンピュータ、ディープブルーが勝利したのは1997年である。

一方日本では「チェスはコンピュータが勝っても将棋は勝てない」という意見が多かった。しかしながら、実際には2010年には清水女流王将（当時）に勝利し、2013年には第2回将棋電王戦でA級棋士を含んだ5名の棋士との対戦で、3勝1敗1引き分けとコンピュータが勝ち越した。その後2015年に情報処理学会が将棋におけるコンピュータの勝利を宣言している。

ちなみに、将棋で人間がコンピュータに負けてもなお「囲碁は難しいから向こう10年は人間が勝つ」等と言われていたが、2016年にはGoogleが作ったアルファ碁が世界最強レベルのイ・セドル

九段に勝利したことで、囲碁においてもコンピュータが人間を打ち負かしたと言って良いだろう。

チェス、将棋、囲碁といったゲームは、完全情報ゲームと呼ばれる種類のゲームである。すなわち、すべての情報が完全に公開されているゲームである。完全情報ゲームの難しさを表1に示す。特に、囲碁はその複雑さが膨大で、その大きさは宇宙にたとえられていた。その複雑さ故にコンピュータにはすべてを把握することが出来ず、人間による直感が勝るなどまことしやかに言われていたが、アルファ碁が登場してわかったのは、なんのことはない、人間は囲碁の宇宙のほんの一部しか理解しておらず、アルファ碁の方が人間より囲碁の宇宙の探索が進んでいたということである。このようなゲームの中でも最も難しいものの一つと言われていた囲碁でコンピュータが人間に勝ったことで、この種のゲームにおいて人間がコンピュータに勝つ見込みはなくなったと言えよう。

表1 完全情報ゲームの難しさ (https://en.wikipedia.org/wiki/Game_complexityより筆者が作成)

| Games | Tree Complexity |
|------------|-----------------|
| リバーシ (オセロ) | 10^{58} |
| チェス | 10^{123} |
| 将棋 | 10^{226} |
| 囲碁 | 10^{360} |

完全情報ゲームと不完全情報ゲーム

人工知能の方が人間より強い、と断言されると、まだ人間が負けていない部分があるはずだ、と言いたくなるのが人情である。

実は、ゲームの世界には完全情報ゲームだけではなく、不完全情報ゲームと呼ばれるタイプのゲームが存在する。不完全情報ゲームとは、情報が完全には公開されていない種類のゲームである。

チェスなどのゲームでは、最適手が存在し、計算時間と計算リソースを無限にとることが出来れば唯一の解が決まるはずである。究極的に言えば、将棋で振り駒をすればその瞬間に勝敗が決まる。一方、不完全情報ゲームでは情報が完全には公開されていないため、最適手がどれかを決定づけることが出来ない。そのため、相手の手を推測しながら自分の手を決めていくという作業が必要となり、完全情報ゲームよりも難しいゲームとなる。

不完全情報ゲームの中で日本人にもなじみ深いゲームの代表が、麻雀であろう。麻雀では、相手の手がどのようなものかわからないため、リーチをかけられれば相手の上がり牌を予測しながら自分の手を進めていかなければならなくなる。このような不完全情報ゲームはまだコンピュータが人間を完全に打ち負かすには至っていない。そのため、新たに人工知能と人間を戦わせる舞台として、不完全情報ゲームに注目が集まっている。

人狼知能

不完全情報ゲームの一つである人狼ゲームを人工知能にプレイさせようという試みが、筆者らが進めている人狼知能プロジェクトである(鳥海・他, 2016)。人

狼ゲームは、ロシアやアメリカで遊ばれていた「村人に紛れ込んだ人狼を探し出す」タイプのゲームの総称である。人狼ゲームのルールなど詳細については片上・他(2015)をご参照いただきたい。

人狼ゲームはよく「嘘をつくゲームである」といわれるが、実態はそうではない。相手を信頼するかどうかを判断するゲームであり、相手を信頼させるゲームでもある。もし人間と人工知能がお互いに嘘をつく可能性があるとなると、自分が持つ情報がいかに正しいかを説得する必要が出てくる。すなわち、相手の言葉に嘘があるという前提に立った時点で、たとえ本当の情報を伝える場合でも、それが嘘ではないことを相手に納得させる材料が必要になるのである。その意味では、人狼ゲームにおいて人工知能が人間を打ち負かす日というのは、人工知能が嘘をつくかもしれないという前提を持った上で、なお人間が人工知能を信頼した時であると言えよう。

では、人工知能はどうやって人間から信頼を勝ち取れるのだろうか？ それを明らかにするためには、なぜ人間が他の人を信頼するのかについて明らかにする必要があるだろう。長期的な信頼関係は、繰返しや互惠性によって構築されるという話がゲーム理論によってなされることがあるが、短期的な信頼関係はどうだろうか？ 人狼ゲームにおいては高々数分のうちに誰を信頼して誰を信頼しないかを決めていくのだが、そこに何らかの根拠はあるのだろうか。

さらに、人狼というゲームは単に勝ち負けを競うというよりも、ともに楽しむという要素が占める割合が大きいゲームであるため、人狼をプレイする人工知能はゲームに強いことよりもゲームに面白

さを提供するべきであろう。では人間を楽しませるためにはどうすれば良いのだろうか？ どのようなプレイがあれば人間は人工知能とのプレイを楽しかったと感じるのだろうか。

「人間の信頼を得る」「人間を楽しませる」知能を実現するためには、将棋や囲碁のように「強ければ良い」というゲームとは違う、新たな人間との接し方を考慮した知能が必要となる。ゲームにおける人間vs人工知能は、単に強いだけの人工知能を作るという目標から、人間とともにゲームをするという新たなステージに入っていくと考えると良いだろう。

このような知能を実現するためには、人間の心理を理解することが大切である。もし、心理学ワールドをお読みの方で、この原稿を「信頼」し協力関係を結んでも良いと考えた方がいれば、是非人狼知能プロジェクトにご参加いただき、新しい人工知能vs人間の形を探っていただければと思う。

文 献

- 片上大輔・他(2015)人狼知能プロジェクト.『人工知能』30, 65-73.
Shannon, C. (1950) Programming a computer for playing chess. *Philosophical Magazine*, 41, 256-275.
鳥海不二夫・他(2016)『人狼知能：だます・見破る・説得する人工知能』森北出版
Turing, A. M. (1953) Digital computers applied to games. B. V. Bowden (ed.) *Faster than thought: A symposium on digital computing machines*. <http://www.turingarchive.org/browse.php/B/7>