



Michael Bowling et al. : Heads-up Limit Hold'em Poker is Solved

Science, Vol.347, Issue 6218, pp.145-149 (2015)

新連載について

今月号から「有名論文ナナメ読み」と題する本コーナーが始まることになった。毎月の特集とは異なる角度から、コンスタントにアカデミックな情報を読者にお届けしたい。

本コーナーの趣旨は、情報処理に関する論文の中で、有名な論文や話題を呼んだ論文を、できるかぎり分かりやすく読者に紹介することである。分量は見開き2ページとし、5分(?)で読めて、少なくとも読者の雑談ネタになる程度にはその研究の面白さをお伝えすることを目指している。もちろん、2ページで論文の内容を細かく紹介するのは無理なので、論文の内容そのものというよりは、研究の背景や論文の重要性、手法のエッセンスをお伝えするのが具体的な目標である。

不完全情報ゲーム

記念すべき第1回で取り上げる論文は、Michael Bowling らによる“Hheads-up Limit Hold'em Poker is Solved”と題された、2015年にScienceに掲載されたポーカーに関する論文である。ポーカーが「解けた」と称する論文タイトルは大胆だ。

2016年Google DeepMindによる囲碁プログラムのAlphaGoが人間のトッププレイヤーに勝利して世界的なニュースになったように、ゲームはAI研究のプラットフォームとしてよく使われる。囲碁や将

棋が、プレイヤーがゲームの状態を完全に知ることが可能な「完全情報ゲーム」であるのに対して、ポーカーは相手プレイヤーの手札が見えないため「不完全情報ゲーム」になっている。不完全情報ゲームは、完全情報ゲームよりも複雑で現実の問題設定に近く、AI研究の大きなターゲットとなっている。

図-1に簡単な不完全情報ゲームであるKuhn Pokerのゲーム木を示す。丸印がゲームの各状態を表し、印の中の文字がその状態でアクションを行うプレイヤーを示す。AとBは各々のプレイヤーを表し、Cは確率的な事象を表現するためのノードを示す。四角はゲームの終了状態を表し、その中にある数字は、プレイヤーAから見た利得を表している。図中で、プレイヤーAのノードが橙色の太い線で結ばれているが、これは、プレイヤーAが実際にゲームをプレイする際には（情報が不完全なため）結ばれたノー

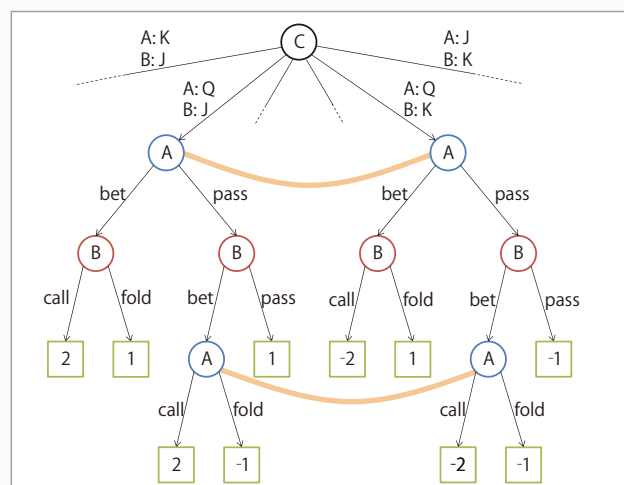


図-1 展開型ゲーム木と情報集合 (Bowling (2015) の図1を元に作成)

ドの間の区別ができないことを示している。このようなノードの集合を情報集合と呼び、プレイヤーは各情報集合で同一の行動戦略(アクションの確率分布)をとらなくてはならない、という制約があるのが不完全情報ゲームの大きな特徴である。そのため、完全情報ゲームでよく用いられるミニマックス探索やモンテカルロ木探索のような手法をそのまま適用することはできない。

ポーカー

ポーカーにはさまざまな種類があるが、この論文が対象としているのは、欧米で人気の「テキサス・ホールデム (Texas hold'em)」と呼ばれるポーカーである。テキサス・ホールデムは、通常2人～10人のプレイヤーで行われ、各プレイヤーは2枚の手札と、コミュニティカードと呼ばれるプレイヤー全員で共有する5枚のカードを組み合わせて役を作る。コミュニティカードは、ラウンドごとに徐々に明らかになっていくため、その過程で、自分が高い役を作れそうであれば掛け金を吊り上げ、勝てそうもなければ早々に勝負から降りるとというのが基本的な戦略である。ポーカーが面白いのは、たとえ自分が高い役を作ることができなくても、掛け金を吊り上げることで相手が怖がって降りてくれれば勝てるということである。そのため、「ブラフ」や「はったり」と呼ばれる戦術が本質的に重要な役割を果たす。

ポーカーを含め、対戦ゲームの(ある種の)最善の戦略は、ゲームの「ナッシュ均衡」を計算することで得ることができる。ナッシュ均衡とは、どのプレイヤーも、自分の戦略を変更することによって得をすることができない状態、すなわち、各プレイヤーの戦略がほかのプレイヤーの戦略に対する最適応答戦略になっている状況であり、それを達成する戦略はそのゲームの「最善」の戦略となる(ただしプレイヤーが3人以上で、「弱い」プレイヤーが含まれる場合には必ずしもその限りではない)。

テキサス・ホールデムのような大規模な不完全情報ゲームのナッシュ均衡解を計算することは、計算量的な問題から長らく困難とされてきたが、CFR (Counterfactual Regret Minimization) と呼ばれるナッシュ均衡の近似計算手法が登場したことにより、一気に解けるゲームの規模が拡大した。CFRでは、図-1に示したようなゲーム木を繰り返し走査することによってそのゲームのナッシュ均衡戦略を近似計算することができる。

Bowlingらはこの論文で、CFRをさらに改良したCFR+と呼ばれるアルゴリズムを用いて、Heads-up (プレイヤー数が2人)、Limit (掛け金の上げ幅が固定) という設定のテキサス・ホールデムのナッシュ均衡戦略を求めることに成功している。彼らがこの計算に使った計算リソースは4,800個のCPUコアを68日間使用という驚きの規模である。

CFR+はあくまでも近似計算の手法であり、厳密なナッシュ均衡戦略が求められたわけではないが、彼らが到達目標とした近似の基準が面白い。彼らが求めた戦略の近似精度は0.986mbb/gというもので、これは、1時間あたり200ゲームを毎日12時間、一生の間(70年間)、完璧な戦略を持つ対戦相手と勝負しても統計的に有意に負け越すことはないという水準である。

彼らが算出した戦略に従ってプレイすればたとえ神様が相手でも負けないというわけだが、きわめて人間的な「はったり」のような行為ですらそれを選択すべき確率がコンピュータに計算されてしまうというのはなかなか切ないものである。

(2017年4月4日受付)

.....

鶴岡慶雅 (正会員) tsuruoka@logos.t.u-tokyo.ac.jp

2002年東京大学大学院博士課程修了。博士(工学)。マンチェスター大学研究員などを経て、2009年北陸先端科学技術大学院大学准教授。2011年より東京大学大学院准教授。自然言語処理、ゲームAI等に関する研究に従事。