数理の窓

負けないポーカープログラム

今年1月、米科学誌サイエンスの電子版に「ヘッズアップ・リミット・テキサス・ホールデムが解析された」との記事が掲載された。カナダのアルバータ大学の研究チームは、人間が一生かけても勝てないコンピュータプログラムを完成したという[1)]。

「テキサス・ホールデム」は、カードゲームのポーカーの一種である。プレイヤーは自分だけが知る手札カード2枚と、ゲームの進行に伴い3枚から最大5枚まで開かれ全員が見られる共通カードから、5枚を選んでポーカーの役を作り、その強さで勝敗を決めるゲームである。ゲームには最大4回の賭けラウンドがある。全員に2枚の手札が配られた後、3枚の共通カードが公開された後、4枚目の共通カードが追加された後、5枚目が追加された後である。プレイヤーはラウンド毎に、①賭け金を他のプレイヤーに揃える、②賭け金をさらに上げる、③ゲームから降りる、の3種類の行動から戦略を作っていく。

テキサス・ホールデムは相手カードが分からない中、最適な意思決定を下す不完全情報ゲームとして研究されてきた。今回の「ヘッズアップ・リミット」は、プレイヤーが2人だけで、各ラウンドの最大賭け回数と毎回の賭け金が制限されている形式であり、テキサス・ホールデムの中でも比較的シンプルなパターンである。それにしても、ゲームを解析するには、1プレイヤーが直面しうる3.19×10^{14}種類もある意思決定の局面に対して、最適な戦略を算出する必要がある。しかも相手カードが分からないので、不確実な現状を推定しながら計算することになる。

今回、この天文学級の処理が実現できたのは、CPUパワーと、「CFR+」[2)]という過去の行動に対する後悔を測って次に取り得る行動の確率を更新していく自己学習型アルゴリズムのおかげである。研究チームは数千個のCPUを用いて、「CFR+」アルゴリズムでプログラムを自己対戦させた。この自己対戦中の学習を通じて、プログラムは後悔を徐々にゼロへと収束させ、負けない戦略を作り出した。

金融機関も不完全情報の局面に対して様々な意思決定をしなければいけない。今後このような大規模計算テクノロジーと最適化アルゴリズムは金融機関でもきっと役に立つだろう。

ところで、アルバータ大学のウェブサイト[3)]では今回のプログラムと対戦できる。そこで自己学習して行けば、負けないポーカープレイヤーになれるかも・・・・　　（朱 映奇）

1）常に勝つという意味ではなく、連続プレイしたら理論上負けないという意味である。
2）Counter Factual Regret minimizationの略。2人ゲームの場合、ナッシュ均衡戦略へ収束するアルゴリズムである。
3）http://poker.srv.ualberta.ca/

野村総合研究所　金融ITナビゲーション推進部　©2015 Nomura Research Institute, Ltd. All rights reserved.