連載

5分で分かる!? 有名論文ナナメ読み

美添一樹（理化学研究所）

Silver, D. et al. : Mastering the Game of Go without Human Knowledge

Nature Vol.550, pp.354-359（19 Oct. 2017）

AIと囲碁

元世界ランク1位の囲碁棋士イ・セドルにAlphaGoが勝利したのは2016年3月のことである．2014年時点ではまだプロ棋士より三段以上弱かった囲碁プログラムが2年足らずの間に急速に進歩を見せた理由は，深層学習，ゲーム木探索および強化学習の3つの技術である．さらにその1年半後には，人間の棋譜を用いずに囲碁のルールだけから自己対戦を繰り返して学習したプログラムがさらなる強さを獲得した．この論文はそのプログラムAlphaGo Zeroについて述べたものである．

これらの成果は，背景となる理論や技術以外にもGoogle DeepMind社が投入した資金，人的資源の豊富さが大きく貢献している．そもそもなぜ，AI研究をリードする企業が囲碁というゲームに労力を投入したのだろうか．それは二人零和有限確定完全情報ゲームに分類されるゲームの多くではコンピュータが人間を上回っている中で，囲碁だけが最後まで人間優勢だったからだと思われる．

コンピュータにとっての囲碁の難しさは局面評価の難しさだったといってよい．黒と白のどちらがどれだけ優勢なのか，計算機上で数値化することは2015年に深層学習の応用が成功するまで不可能であった．

画像認識と囲碁

猫画像を認識するという成果に象徴されるように深層学習は多くの分野で進歩をもたらしている．囲碁の盤面は画像と見なすことができる．そこで画像認識で培われた深層畳み込みニューラルネットワーク（Deep Convolutional Neural Network, DCNN）を囲碁に使うのは自然な発想である．

画像認識での入力は赤緑青の3枚の画像として与えられるのが一般的である．同様に，囲碁の盤面から何枚かの画像を生成し，それを入力としてDCNNに与えることが考えられる．AlphaGo ZeroのDCNNは入力から**着手予測**および**局面評価**の2つを出力する．

着手予測は犬・猫などを予測する分類問題と同様だといえる．画像認識で近年標準的に用いられるImageNetベンチマークでは1,000種類の物体の分類を行わせる．この場合は入力はRGB画像で，出力は1,000個の重みとなる．AlphaGo Zeroの着手予測では，図-1に示すように現在の盤面や過去の手の履歴から生成した17枚の画像を入力とする．出力は，碁盤が19路なので着手数に合わせて19×19＝361通り（パスを加えて362通り）の重み$\boldsymbol{p}$となる．さらに局面評価は同様の入力から[−1, 1]の数値vを出力する（1なら黒勝ち−1なら白勝ち，0なら互角）．画像認識で高い精度を達成するResidual Network（ResNet）をそのまま用いて人間の棋譜を学習すると，高段者の着手を60%以上の精度で予測できる．

ゲーム木探索

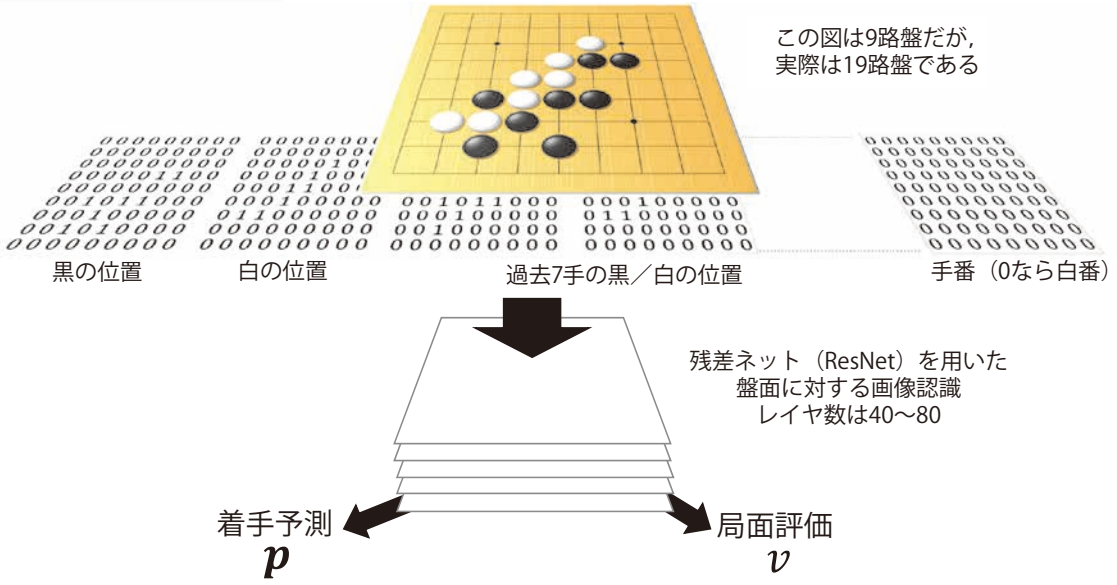
しかし，高い精度で着手を予測するだけではプロ棋士にはまったく及ばない．正しい着手を選ぶには**ゲーム木探索**を行う必要がある．すべての手を読むことは現実的な時間では不可能なため，着手予測と局面評価に基づいて枝刈りをして先読みを行う．探索アルゴリズムとしてはコンピュータ囲碁で標準的に用いられている**モンテカルロ木探索**が用いられる．

多くの探索アルゴリズムは局面に対して単純な評価値があるとして探索を行うが，モンテカルロ木探索は局面に勝率があるとして探索を行う．AlphaGo は実際にサンプリング（ロールアウト，プレイアウトなどと呼ばれる）を行っていたが AlphaGo Zero では DCNN の出力する局面評価を擬似的な勝率と見なして探索を行う．

ゼロからの強化学習

ある状態から行動を起こし，得られる報酬を手がかりに学習を行う方法が強化学習である．古くから知られている手法だが，深層学習と組み合わせた深層強化学習（Deep Reinforcement Learning）がビデオゲームのスコアを報酬とした学習で成功するなど，注目されている．囲碁では報酬は勝敗で得られる．手数が多く，報酬が得られるのがかなり先になるため学習は困難なはずだが，深層強化学習は囲碁でも成功した．囲碁のルールだけを与え，自己対戦から DCNN のパラメータを学習することによって作成されたプログラムが AlphaGo Zero である．

まず DCNN のパラメータをランダムに初期化して自己対戦を行い，対局途中で勝敗と（探索結果の）着手をより正確に予測可能なようにパラメータを更新し，更新されたパラメータを用いたプログラムの方が強くなっていれば採用する．大規模な計算資源を用いてこの手順を 40 日間繰り返し，最終的には冒頭で述べたイ・セドルに勝利した AlphaGo に対して単純な期待勝率が 99.9% を超える強さに達した．人類を超越した強さを持ち，そのために当然ながら人間の着手の予測精度は下がる．まさにゼロから学習したため，囲碁棋士の知る定石とは異なる手を打つ．アルゴリズムの性質から生じる弱点があるはず



◆図 -1　AlphaGo Zero の DCNN

だが，プロ棋士であっても意図的に弱点を狙うことは大変困難である．

別論文だが同じ方法でチェスと将棋も過去最高のプログラムを上回った．人間の棋譜からの学習をまったく用いずに，複数のゲームで人類を圧倒する強さを達成したことは大きな意義がある．

AlphaGo Zero の学習で使われた計算資源は論文に明示されていないため推測するほかないのだが，行われた計算の量と非公式な発言などを手がかりとした筆者の推測では TPU を 2,000 台[☆1] 程度用いたと思われる．個人所有のゲーム用 PC で同等の計算を行うには 100 年単位の時間がかかると推測される．しかし規模を縮小しての再現実験は可能である．企業や個人によって作成された AlphaGo Zero クローンといえる囲碁プログラムはすでに多数あり，オープンソースソフトウェアとして公開されているものも複数ある．手軽にプロ以上に強い囲碁プログラムと対戦できる環境が整っている．

囲碁界が受けた衝撃は非常に大きく，世界中のプロ囲碁棋士が盛んに検討を行った結果として AlphaGo の着手の解説を試みた書籍がすでに多数出版されている．最近はプロ棋士同士の対局でも新しい発想の着手が頻繁に見られ，一囲碁ファンとしては観戦していて非常に面白い．同時に，以前は夢物語であった考えをもてあそぶことが増えた．仮に自分の専門分野で自分を遙かに上回る AI が出現した際に，果たしてプロ棋士と同じように真摯に AI から学ぶ姿勢を持てるだろうか．

(2018 年 6 月 1 日受付)

☆1 Google 社が開発した Tensor Processing Unit ver.2 を 4 個搭載した計算機を 2,000 台用いたという推測．

美添一樹（正会員）kazuki.yoshizoe@riken.jp

2009 年東京大学大学院より博士号取得．（株）富士通研究所，東京大学研究員，同助教などを経て，2017 年より理化学研究所革新知能統合研究センター，探索と並列計算ユニットのユニットリーダー．探索アルゴリズム，並列計算，ゲーム AI 等に関する研究に従事．