

Article AlphaGo の技術と対戦

Technology and Games of AlphaGo

伊藤 毅志

Takeshi Ito

電気通信大学

The University of Electro-Communications.

ito@cs.uec.ac.jp, <http://minerva.cs.uec.ac.jp/~itolab-web/wiki.cgi>

松原 仁

Hitoshi Matsubara

公立はこだて未来大学

Future University Hakodate.

matsubar@fun.ac.jp

Keywords: computer go, deep learning, Monte-Carlo tree search, reinforcement learning.

1. ゲーム情報学から見た囲碁

囲碁は、ゲーム情報学的に見ると、“二人零和有限完全情報確定ゲーム”と分類される。これは、二人でプレイする、一方が勝てばもう一方は負ける（あるいは引き分けになる）、ルール上いつか勝負がつく、相互にお互いの合法手（ルール上許される手）がすべて明らかにされている、サイコロのような不確定な要素を含まないゲームであるということの意味する。このようなゲームは、世界中にたくさんあり、チェス、将棋、中国象棋のようなチェスライクゲームだけでなく、オセロやチェッカーなどのゲームも存在する。ゲームの複雑さを、平均合法手と平均終了手数（初手からゲーム終了までの平均手数）から探索量という単位で計算すると、比較的プレーヤ人口の多いゲームは、以下のように概算される（例えば将棋は平均合法手が約 80 で平均終了手数が約 115 なので、80 の 115 乗ということではほぼ 10 の 220 乗になる）。

表 1 ゲームとその探索量

| ゲーム | 探索量 |
|-------|------------|
| チェッカー | 10^{30} |
| オセロ | 10^{60} |
| チェス | 10^{120} |
| 将棋 | 10^{220} |
| 囲碁 | 10^{360} |

この表を見ると、囲碁は、この種のゲームの中で群を抜いて複雑なゲームであるといえる。チェッカーは、2007 年にカナダの研究グループが完全解（先後が最善を尽くすと引き分けになる）

を示しており [Shaeffer 07]、オセロ、チェスは、1997 年に人間のトップに勝利している [Buro 97, Pandolfi 98]。将棋は、2015 年に情報処理学会が人間のトップに勝つコンピュータ将棋プロジェクトの終了宣言を行っている [松原 15]。しかし、囲碁だけは、2015 年の時点で、人間のトップにはほど遠く、まだアマチュア 6 段程度であった。これが、囲碁がゲーム AI 研究の最後の砦といわれてきた所以である。

チェスや将棋などのゲームは、ゲーム木探索と呼ばれる手法で強い AI を実現してきたという歴史があるが、囲碁は、他のゲームに比べて圧倒的に平均合法手が多いだけでなく、ゲーム木探索の末端ノードで必要とされる評価関数（先後手どちらがどれくらい優勢なのかを数値的に表現する関数）の設計が絶望的に難しいとされており、ゲーム木探索の手法が全く使えないというネックがあった。

コンピュータ囲碁の歴史を見ると、1960 年代には研究が始められ、1970 年代には局面を評価するためのさまざまな手法が考えられ、開発者のもっている囲碁の知識を総動員して、何とかコンピュータ上で表現しようと試みられた。その後、1980～90 年代には、ニューラルネットワークや組合せゲーム理論、認知科学的アプローチなどさまざまな情報処理技術の適用が試みられたが、2006 年にモンテカルロ木探索 (MCTS) の手法が現れるまでは、長い冬の時代が続いていた。アマチュア初段も厳しいほどであった。

2006 年に現れた MCTS は、それまでの手法を一掃する画期的な手法で、コンピュータ囲碁はそれまでの遅れを取り戻す勢いで、急激な進歩を遂げた [美添 12]。2008 年に初めてプロ棋士を相手に 9 子局（九つの石をあらかじめ置いて始めるハンデ戦）で勝利を取めると、その年のうちに 8 子局、7 子局でも勝利を取ってしまう。2011 年には 6 子局、2012 年には 5 子、4 子で次々とプロ棋士を破った。その後、2013～15 年の間は 4 子で勝ったり負けたりを繰り返す状態が続き、MCTS による手法も頭打ちを見せ始めており、次のブレークスルーが求められていた。

2. AlphaGo の発表と技術

そんな中、2016 年 1 月 28 日に、ディープラーニングを用いたコンピュータ囲碁の人工知能に関する発表が世界を駆け巡った。Google の研究グループが Nature にコンピュータ囲碁に関する論文を発表し [Silver 16]、記者会見を行ったのだ。それによると、ディープラーニングと強化学習を組み合わせた新しい手法を用いて、既存のコンピュータ囲碁プログラムに対して 99.8% の勝率を実現し、さらにヨーロッパチャンピオンのプロ棋士（二段）に対して、互先（ハンデなしの対局）で 5 連勝する強さを実現したというものであった。前章にも書いたように、モンテカルロ木探索の手法はやや行詰まりを見せていただけに、コンピュータ囲碁関係者に衝撃が走った。

予兆はあった。2014 年から強い

プレーヤーの棋譜を教師データにしたディープラーニングを用いることで、プロ棋士の手を予測する手法が提案された論文が出始めており、それまでの手法では 40% 程度しか予想できなかったものが、57% の予測率を実現する論文が発表されていた [Clark 14, Maddison 14]。この予測率はかなり高い値で、多くの開発者はこの手法を模倣しようとしていた。しかし、これを実現したハードウェアの構成は、GPU を非常に多く使ったものであり、一般の開発者が容易に模倣できなかった。それでも、多くの開発者は導入を試み始めていた。

そんな中、本年 1 月のあの論文が突然発表された。AlphaGo は、以下の三つの段階から構築されている（以下はあくまで 1 月の Nature の論文の時点での話であり、3 月のイ・セドルの対戦のときにはさらに進んだものになっていた可能性があることに注意されたい）。

〈第 1 段階〉

13 層の Deep Convolutional Neural Network (DCNN) を使い、インターネット囲碁道場 KGS の 6 ~ 9 段（これはかなり強いレベルの人間である）の棋譜 16 万局（約 3 000 万局面）を用いて、教師あり学習 (Supervised Learning) を行い、次の一手の予測器を作成した。ある局面でその人間が打った手を正解として、それを予測するプログラムである。

50 個の GPU で 3 週間かけて学習を行い、これによって予測率 57% の高確率の予測器を実現した。この DCNN は SL Policy Network と呼ばれる。

〈第 2 段階〉

上述の SL Policy Network を初期値とし、「自己対戦によるゲームの勝利」を報酬として強化学習 (Reinforcement Learning) を行った。ゲームにおける強化学習は例えばバックギャモンを対象とした研究が有名である。

第 1 段階のものと対戦させた場合に 80% 以上の勝率をあげる DCNN を実現した。この DCNN は RL Policy Network と呼ばれる。なお、この学習は 1 日で行われた。

〈第 3 段階〉

自動生成したデータから、局面の勝率を計算する DCNN である Value Network を作成した。

はじめに、途中まで SL Policy Network で自己対局させて、一手ランダムに打ち（ここがポイントである）、その後は RL Policy Network に勝負がつくまで打たせるということを繰り返し、3 000 万局面のデータを自動的に生成した。ここでデータといているのは、ランダムに打った直後の局面とその結果としての勝敗である。このデータから教師あり学習を行って、RL Policy Network によって局面の勝率を計算する Value Network を形成する。この学習には 50 個の GPU が用いられて 1 週間が費やされた。この Value Network が事実上の囲碁の評価関数であり、AlphaGo の強さの根源である。

〈第 4 段階〉

このようにして作成された第 1 段階の SL Policy Network と第 3 段階の Value Network と、従来の手法である MCTS を組み合わせて AlphaGo は構成されている。第 2 段階の RL Policy Network は、第 3 段階の Value Network を形成するためにだけに使われており、実際の対局では使われていないのは興味深い。

第 1 段階までは、2014 年までに発表された技術であり、すでにいくつかのコンピュータ囲碁プログラムも規模は違うものの導入が試みられていた。

第 2 段階以降が AlphaGo の新しい技術である。ここが AlphaGo の強さの秘密であるように思われる。特に不可能と思われていた囲碁の盤面評価関数を Value Network という形で勝率という形で表現できるようになったことは驚きである。

これを実現したのは、Google という大企業による潤沢なリソースがあることはいうまでもない。個人レベルで高価な GPU を 50 個もそろえることは容易ではないし、プロ棋士と対戦した AlphaGo は 1 000 個以上の CPU と 176 個もの GPU を使用しており、膨大なマシンパワーが費やされたことは事実である。さらに論文の著者は 20 名にも

上り、これだけの優秀な人材をたくさん集めたこともこの成功を支えている。

3. AlphaGo の達成したことと課題

Nature の時点での AlphaGo は確かにこれまでのコンピュータ囲碁よりも格段に強かったものの、まだ二段という弱いプロ棋士に勝っただけであり、トップレベルのプロ棋士とはかなりの実力差があったと思われる。2016 年 3 月 9 日から 15 日にかけて韓国のソウルで囲碁のプロ棋士のイ・セドルとコンピュータ囲碁 AlphaGo が 5 局互先（ハンディなし）で戦うというニュースが流れたときは、ほとんどのプロ棋士およびコンピュータ囲碁の関係者はイ・セドルが 5 連勝すると思っていた。しかし蓋を開けてみると AlphaGo が 4 勝 1 敗で勝利をした。Nature の時点からはるかに強くなっていたのである。しかもイ・セドルが明らかなミスをして勝ったというのではなく、AlphaGo が実力でイ・セドルを圧倒していた。AlphaGo は序盤でイ・セドルや解説のプロ棋士が理解できない手を打った。人間の常識からすると悪手に見えたのだが、対局が進んでみるとその手は良い手であることがわかった（人間がそれに気付いた時点ですでに手遅れで勝負は決まっていた）。これは、人間には見えない未来がコンピュータだけに見えていたことを意味している。大きな驚きであった。

AlphaGo が実現したことをまとめると、以下のようなことがあげられる。

- 1) DCNN を用いて、膨大な棋譜データからの学習でトッププロ棋士の直観的な手の生成に匹敵するほどの予測器をつくり上げた。
- 2) 自動生成したデータから、囲碁のような複雑な局面の優劣を理解する評価関数のようなものを Value Network という形で実現した。
- 3) 潤沢なハードウェアを用いて従来の MCTS の手法を組み合わせることで、トッププロ棋士に勝ち越すほどの囲碁のパフォーマンスを実現できることを示した。囲碁の局面認識のような複雑な問題

をディープラーニングの手法でうまく実現できることを示したことは、正直、大きな成果であったといえるだろう。これまでにディープラーニングが成功してきたのは画像認識や音声認識というパターン認識の分野であった。意思決定や問題解決という従来は主に記号処理で行ってきた対象についてはそれほどうまくいっていなかった。囲碁という対象はパターン認識と記号処理のちょうど中間に存在すると考えられる。囲碁は局面が(19×19=361か所)非常に広いので、人間も局面をパターンとして認識していると考えられる。パターン認識の結果に基づいて記号処理を(も)行って次の一手を考えるという問題解決を行っている。今回のAlphaGoの成功は、ディープラーニングがパターン認識から記号処理へと適用範囲を広げる走りとなることが期待される。

しかし、AlphaGoが実現できなかった問題も残されている。イ・セドルとの4局目では、思わぬ弱点をさらすこととなった。ここで露呈した問題は、MCTSを使っている以上避けられない問題であることは、実は対戦以前から指摘されていた。局面の直観的な認識能力は格段に上がったが、AlphaGoの探索部分は、相変わらずMCTSを使っている。モンテカルロアプローチの問題点として、探索の過程で乱数を使っている点があげられる。乱数を使って勝率の計算を行う以上、必然手の応酬のような直線的な先読みが困難であるという問題点が残っているのである。AlphaGoが間違えた局面も複雑で正確な先読みが必要とされる局面であった。AlphaGoはこういった複雑で正確な先読みの能力においては、まだプロ棋士には及ばないように見える。

2012年に狭い盤面で争う九路盤の囲碁において、プロ棋士3名と当時の世界チャンピオンプログラムZenが対戦

したが、実は6戦全敗している[伊藤13]。十九路盤よりかなり狭い九路盤では、読みの精度がクリティカルに影響し、MCTSの弱点がより露骨に現れる。正確な読みができるプロ棋士は、コンピュータの弱点を見切ったために、このような一方的な展開になったと思われる。AlphaGoもおそらく同じ弱点を抱えている。

人間のプレーや囲碁という問題を多角的に捉えることができる。局面全体を見て直観的に捉えることもできるし、局面を分割して部分的な石の死活や石の強弱を認識することも可能である。もっと言えば、九路盤という盤の大きさにも対応して、ゲームを認識して問題を解くことができる。AlphaGoは、教えられた枠組みの中で人間を超える直観的認識を身に付けた。ただ、局面を別の視点で切り分けて考えたり、囲碁という問題をどうやって解いたらよいかを自ら導いたりすることはできない。また、DCNNによって獲得した知識をかみ砕いて、人間の理解しやすい形で説明することもできない。

こういった課題は、多くの人工知能がこれから人間の社会に入っていく際の軋轢を生み出す可能性がある。AlphaGoのような囲碁という例題をもとに、人と知的システムとの新しい関係を考えていく契機になるかもしれない。

◇ 参 考 文 献 ◇

- [Buro 97] Buro, M.: The othello match of the year: Takeshi Murakami vs. Logistello, *ICCA J.*, Vol. 20, No. 3, pp. 189-193 (1997)
- [Clark 14] Clark, C. and Storkey, A.: Teaching Deep Convolutional Neural Networks to Play Go, arXiv:1412.3409 (2014)
- [伊藤 13] 伊藤毅志 ほか: ミニ特集「コンピュータ囲碁の最前線〜九路盤囲碁のイベントから〜」, 情報処理, Vol. 54, No. 3, pp. 232-250 (2013)
- [Maddison 14] Maddison, C. J., Huang, A., Sutskever, I. and Silver, D.: Move Evaluation in Go Using Deep

Convolutional Neural Networks, arXiv:1412.6564 (2014)

- [松原 15] 松原 仁: コンピュータ将棋プロジェクトの終了宣言, 情報処理, Vol. 56, No. 11, pp. 1054-1055 (2015)
- [Shaeffer 07] Schaeffer, J., Burch, N., Björnsson, Y., Kishimoto, A., Müller, M., Lake, R., Lu, P. and Sutphen, S.: Checkers is solved, *Science*, Vol. 317, pp. 1518-1522 (2007)
- [Silver 16] Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Driessche, G. v. d., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T. and Hassabis, D.: Mastering the game of go with deep neural networks and tree search, *Nature*, Vol. 529, pp. 445-446 (2016)
- [バンドルフィーニ 98] ブルース・バンドルフィーニ: ディープブルー vs. カスパロフ, 河出書房新社 (1998)
- [美添 12] 美添一樹, 山下宏著, 松原仁編: コンピュータ囲碁—モンテカルロ法の理論と実践, 共立出版 (2012)

2016年4月7日 受理

著 者 紹 介



伊藤 毅志 (正会員)

1994年名古屋大学大学院工学研究科情報工学専攻修士(工学博士)。同年より、電気通信大学助手。2007年より同助教。2010年改組により情報・通信工学専攻助教。ゲームを

題材とした認知科学的研究に従事。コンピュータ囲碁フォーラム(CGF)理事。エンターテイメントと認知科学研究ステーション代表。UEC杯コンピュータ囲碁大会および、電聖戦実行委員長。



松原 仁 (正会員)

1986年東京大学大学院工学系研究科情報工学専攻博士課程修了(工学博士)。同年、通商産業省工業技術院電子技術総合研究所(現産業技術総合研究所)入所。2000年公立はこだて

て未来大学教授。2016年同副理事長。人工知能、ゲーム情報学などに従事。本学会会長。情報処理学会理事。コンピュータ囲碁フォーラム会長。