本書はInterface2021年4月号掲載の16ページに加筆修正を加えたものです.

AI自習ドリル

第4回　強化学習2…人工知能 vs. 人工知能

牧野 浩二，足立 悠

強化学習とは，学習者（コンピュータ）に，全ての途中経過は教えず，ところどころで「良い状態になった」または「悪い状態になった」を教えるだけで，コンピュータが試行錯誤しながら少しずつ上手に行動できるようになる学習方法です.

全てを教えなくとも自ら学んでいくため，人間が思いつきもしなかった動作で物事をうまく解決することがあります．人間を超える能力を獲得することもあるので，今後のAIの主力となることを期待されています.

前回は「エージェント」と呼ばれる強化学習で賢くなる個体（プログラム）が1つしかありませんでしたが，今回はエージェントが2つです．競い合いながら学習します.

なお，強化学習は学習方法の枠組みですので，実装するためのアルゴリズムがいろいろ開発されています．この記事では実装が容易でかつ学習手順が直感的に分かりやすいQラーニングを用います.

1　できること

強化学習は，最終的な「良い状態と悪い状態」を与えておくと，自ら途中経路（経過）を学習して，うまく行動するようになる学習方法です.

囲碁や将棋，トランプなどの対戦型のゲームでは，大抵の場合，勝ち負けがはっきりしています．勝ち負けがはっきりしているということは，良い状態と悪い状態がはっきりしていることになります．そのため，強化学習に向いた問題と言えます．強化学習を対戦ゲームに応用した例として以下があります.

● 囲碁

2016年にDeepMind社（現在はグーグル）が開発したプログラム「AlphaGo」が，プロ囲碁棋士を初めて破り，大きな話題となりました．囲碁は縦横19本ずつの直線が交差する361点に交互に石を打つため，最初の2手だけ考えても先手は361通り，後手は360通りあります．そのため，最初の1手を互いが打ち終わっただけでも361 × 360 = 129960通りの盤面が考えられます.

また，全ての考えうる盤面は2^{361}通りあります．考えうる盤面の数が膨大すぎるため，最適な手を探ることは困難だと考えられていましたが，強化学習の考え方を応用することで人間よりも強くなったといわれています.

図1-1に示すのは同社が提供してくれているウェブ・サイトの画面です．ここには次のように記されています.

この学習ツールは，人間同士が打った231,000局と，AlphaGoと人間が対局した75局の棋譜データを基に，囲碁近代史における6,000種類の布石パターンを分析することができます．AlphaGoと棋士たちの打ち方を比較して，囲碁の奥深さを探索してみましょう.

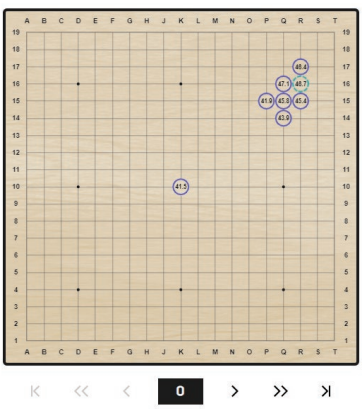

図1-1　強化学習によって囲碁AIは強くなった
AlphaGo学習ツールで新たな囲碁の楽しみを，AlphaGo Tech（https://alphagoteach.deepmind.com/ja）から引用