The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

即時戦略ゲームにおけるAI学習プラットフォームの構築及びDQNによるAIの実装

AI Trainning Platform and AI Implementation with DQN for Realtime Strategy Game

張 翌坤 *1 Zhang Yikun　橋山 智訓 *1 Hashiyama Tomonori　田野 俊一 *1 Tano Shun'ichi

*1電気通信大学 情報理工学部
Dept. of Informatics, The University of Electro-Communications

Research on video game AI has been done for a long time. In recent years, due to the rapid development of neural networks, the game AI which was considered difficult to build has appeared. After these AI beat humans in turn based games with complete information, such as GO, the next research direction has been focused on simultaneous games with incomplete information. The RTS game is one of the simultaneous-move games with incomplete information; therefore, in this paper we report the design of a platform for AI learning in the RTS game whose mechanism has been simplified. In addition, we implemented an AI for this type of game, and carried out some experiments.

1. はじめに

ゲーム AI は、AI 研究の早い段階から研究対象となってきた。コンピュータの性能の上昇と共に、現在では様々なゲームで人間に勝利するようになってきた。特に IBM が開発した DeepBlue[1] が当時のチェス世界チャンピオンに勝利したことは大きなニュースとなった。近年になり DeepMind の AlphaGo[2] が囲碁のプロに勝利し、人工知能研究の成果を示してきた。これらのゲームは完全情報ゲームと呼ばれ、ゲームプレイに関する情報をすべて得ることができる。現在のゲーム AI に関する研究の注目は不完全情報ゲームに移っている。

本研究では不完全情報ゲームのうち即時戦略ゲーム (Realtime Strategy Game, RTS ゲーム) に分類されるオンラインゲームの Clash Royale を対象とする。RTS ゲームの状態数と行動空間はともに囲碁を超えている。Clash Royale を研究対象にすることで、AI の学習目的を RTS ゲームの中の主要目的である対戦に絞ることができる。本研究では Clash Royale の AI 学習環境を構築し、AI を実装したので報告する。

2. 関連研究

2.1 Q Learning

Q 学習は 1992 年に、WATKINS[3] らによって、強化学習の一手法として提案された。ある有限マルコフ決定過程において、すべての状態に対して、必ず最良の行動が存在する、そのすべての状態に対して十分に試行を重ねると、各状態においての各行動の評価値は収束することが証明された。現実問題に対して、連続した状態をデジタル化して、離散的な状態集合に対して学習を行うのが制約となっている。

2.2 Deep Q Learning

DeepMind は、ニューラルネットワークを用いて、Q 関数を再構築した。多層なニューラルネットワークによって構築されたため、Deep Q-Learning(DQN)[4] と名付けられている。この DQN を用いて、Atari ゲーム機で発売されたレトロゲームについて学習させた結果。Breakout、Enduro、Pong などのゲームで人間レベルのプレイ技術に達していた。

2.3 AlphaZero

囲碁のような状態数が多いゲームに対して、単純な先読みは能力不足になる。DeepMind は強化学習の手法を用いて、AlphaGo[2] を開発した。囲碁の盤面の特徴を抽出して、特徴に合わせて方策を決めるネットワークと価値を決めるネットワークを別々に学習した。実行時には両方利用して、人間の棋士に勝利した。しかし、リアルタイムに進行する不完全情報ゲームに対して、応用と検証する例はまだ少ない。

3. AI 構築用ゲーム環境の実装

本研究で構築された疑似ゲーム環境は OpenAI gym のゲーム環境を参考にした。ゲーム内全オブジェクトの処理はフレーム単位で処理される。一フレーム毎に各ユニットのやるべきことを逐次実行する。ゲーム内ユニットの活動の更新が完了後、この瞬間のゲーム内の全情報 (observation) を抽出でき、AI エージェントが取るアクションを入力 (step(action)) できる。ゲーム終了後、両側プレイヤーが獲得した合計リワード (total_reward) が取得できる。構築したゲーム環境は図 1 に示す。

図 1: ゲーム環境の実行の様子

プレイヤーは手持ちの 8 枚のカードのうちの 4 枚のカードを使って、ユニットを自分の陣地に配置する。配置されたユニットは決められた行動で相手を攻撃しに行く。ゲームの目的は相手のタワーを破壊することである。ゲーム終了時に、より多くの相手のタワーを破壊した方が勝ちとなる。

4. AI の構造

今回提案する AI はニューラルネットワークによって実現した。構造は図 2 のように設計した。各層の活性化関数は LeakyReLU を利用した。

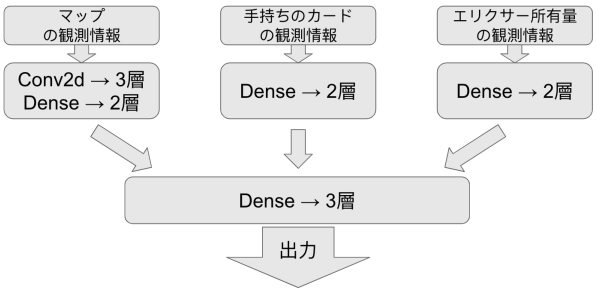

図 2: ニューラルネットワークの構造

5. 実験と結果

今回は AI として、DQN を用いる。ゲーム終了後の合計リワードは次式 1 で計算される。

$$Total\_reward = Time\_bonus * (Reward - Penalty) \quad (1)$$

Time_bonus はゲームの進行に連れて減少する係数であり、ゲームが長引かないようするためのものである。Reward と Penalty それぞれ敵陣に与えだダメージの量と自陣が受けたダメージの量がゲーム勝敗条件と計算した数となる。対戦双方は同じネットワークで判断を下し、学習していく。1 ゲームを 1 エピソードとして、学習は 3000 エピソード行った。対戦両方が勝ったときの合計リワードの 30 期間の移動平均線を図 3 に示す。横軸は訓練のエピソード数、縦軸はゲーム終了後の合計リワードになる。初期に学習は急速に進むが 700 エピソード以降は、合計リワードがほぼ変化していない。

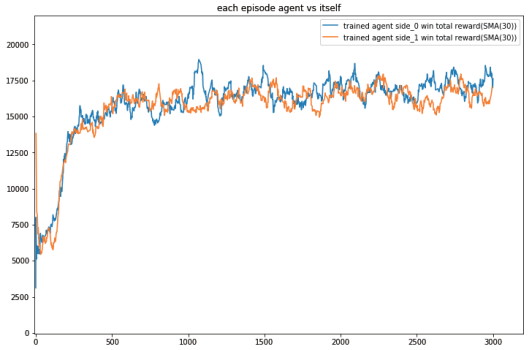

図 3: エピソードごとの自己対戦

ランダムエージェントの自己対戦の結果と学習 1000/2000/3000 エピソードのエージェントとランダムエージェントの対戦結果の合計リワードとゲームの終了までの時間の関係を図 4 に示す。ランダム/シンプル/1000/2000/3000 エピソードエージェントとランダムエージェントの対戦勝敗数を表 1 に示す。

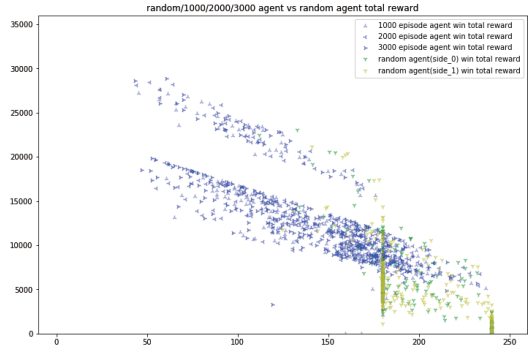

図 4: シンプル/1000/2000/3000 エピソードエージェント対ランダムエージェントの合計リワード

表 1: ランダム/シンプル/1000/2000/3000 エピソードエージェントとランダムエージェントの対戦勝敗数

エージェント	勝利数	敗北数	引き分け数	勝率
random	215	230	55	43%
1000	320	171	9	64%
2000	305	188	7	61%
3000	300	196	4	60%
simple	330	164	6	66%

図 4 で確認できるのは、ランダムエージェント (緑マーク) はゲームの終了時間は約 180 秒後であり、合計リワードは低かった。学習後のエージェント (青マーク) はゲームが早く決着するように学習できた。全体にゲームの終了時間が前倒しになったことが確認できた。ゲーム終了時のエージェントが獲得できたリワードも上昇したことが確認できた。

表 1 から、ランダムエージェントの勝率 43%と低い。学習後のエージェントのランダムエージェントに対する勝率は全体的に 20%上昇し、60%以上の勝率になった。

6. 今後の課題

今回は Clash Royale と類似した RTS ゲームの AI 学習用のプラットフォームを作成した。このプラットフォームで RTS ゲームに関する AI の研究がより簡単にできるようになった。また、今回は RTS ゲームの戦略を学習する AI を DQN により構築した。しかし、700 エピソード以降学習が余り進んでいないと思われる。これからはより勝率の高い AI を学習させるために色々な工夫を考えたい。

参考文献

[1] M. Campbell, A.J. Hoane Jr, H. Hsu, *Deep Blue*, pp.5783, Artificial Intelligence, Vol. 134, Issues 1-2, pp. 57-83, 2002

[2] D. Silver, et. al, *Mastering the game of Go with deep neural networks and tree search* , Nature, Vol. 529, pp. 484489

[3] CJCH Watkins, P Dayan, *Q-Learning* , Springer, Vol. 8, pp. 297292, (1992)

[4] Bellemare, M. G., Veness, J. & Bowling, M. *Investigating contingency awareness using Atari 2600 games.* , Proc. Conf. AAAI. Artif. Intell. 864871 (2012)