深層強化学習に基づく トレーディングカードゲーム環境の構築

第 1 グループ　西村 昭賢

1. はじめに

近年, 人工知能に関する研究分野は目覚ましい発展を遂げており様々な分野に応用されている. その中でも人間の学習プロセスに近いとされる強化学習と深層学習を融合した深層強化学習は実世界の問題解決への応用だけでなく, ゲームにも盛んに応用している. 特に将棋や囲碁といった, プレイヤーが意思決定をする段階でそれ以前の意思決定の過程がすべて把握可能な完全情報ゲームへの応用においては成果が顕著である. 最近では麻雀やポーカーのような, プレイヤーに与えられる情報が部分的である不完全情報ゲームへの応用も注目されている. 本研究では不完全情報ゲームであるトレーディングカードゲーム環境への深層強化学習の適用と深層強化学習を用いたゲームバランス調整手法を提案し, 独自に構築したトレーディングカードゲーム環境を用いて数値実験することでその有効性を示す.

2. 要素技術

2.1. OpenAI Gym

OpenAI Gym [1] は非営利企業 OpenAI が提供する強化学習のシミュレーション用ライブラリであり, 強化学習の環境として多くのゲーム, シミュレータが登録されている. さらには提供されているインターフェースに沿うことで自作の強化学習環境を構築し利用することもできる.

2.2. Deep Q Network

Deep Q Network (DQN) [2] は代表的な価値ベースの強化学習手法である Q 学習に対して深層学習を融合させた深層強化学習手法である. DQN では, ある状態における行動ごとの Q 値を推定することでたとえ状態が連続値であっても学習可能としている. Experience Replay や Fixed Target Network といった工夫により Q 値の更新の過程に深層学習を用いても安定した学習を可能としている.

2.3. Genetic Algorithm

Genetic Alogorithm (GA) とは, 生物の進化と進化の過程を模した最適化手法である. GA では 1 つの解を 1 つの個体として表現し, 多数の個体からなる個体群を用いて解空間の多点を同時に探索する. 各個体はどの程度良い解であるかという指標として適用度を持ち, 個体群に対して選択, 交叉, 突然変異と呼ばれる 3 種類の遺伝演算子を適用させ探索を進めることで最良の適用度を持つ個体を見つける.

3. 提案手法

深層強化学習によりバランス調整を施すデッキにおける最適な戦略を構築したエージェントを作成し, 作成したエージェント同士で先攻, 後攻でカードを 1 枚づつ抜いて対戦していくことで構築された戦略におけるデッキ内のカードパワーを定量的に測定する. その後, 得た結果を基にデッキを環境に取り込む際に用いる GA の解空間の次元を削減しパラメータの変更されるカード枚数を最小限にしてトレーディングカードゲーム環境のゲームバランスを調整する.

4. 実験

4.1. 実験 1

本実験では, カードゲームの対戦環境において深層強化学習を用いて後攻プレイヤーの行動を学習し, 10000 回対戦を実行する. そこで学習済エージェントの勝率, 学習中の獲得報酬の推移を記録し深層強化学習が適用できるかどうか確認する. また対戦した記録から選択した行動, 各カードのプレイされた回数を計測し学習済みエージェントの行動を分析する. 結果として, 勝率は 0.71365 (ここまだデータとってない) とベースラインと比較して (データまだ) 高い勝率を残した. また, エージェントは盤面のカードを積極的に相手に攻撃して早く相手プレイヤーの HP を 0 にする戦略をとっていることが分かった. また, 構築した戦略におけるカードの強弱も学習していることが分かる (card record まだ).

4.2. 実験 2

実験 1 で作成したエージェントを先攻後攻両方に配置し, それぞれデッキからカードを 1 種類ずつ抜いて 10000 回ゲームを実行して先攻側の勝率を記録した. 手動で作成した AI について同様に実験をし結果を比較したところ, デッキ内で最も強いとされたカードは同じものであった一方で, 最も弱いとされたカードは異なっていた. これはエージェントが学習において戦略におけるカードの強弱を学んでいることから単純な戦略な AI で最も弱いと判断されたカードはそもそもゲームに登場することが少ないためと考えられる. 深層強化学習を用いたシミュレーションにおいてカードのパラメータだけでは一見して弱いと判断されずらいカードを検出することが分かった.

4.3. 実験 3

実験 2 のデータからデッキ内のカードについて調整する優先順位を作成し, 調整する枚数を増やしていきながら, 既存の環境においてどの相手に対しても勝率が $50 \pm 5\%$ を残すようにデッキを調整する.

5. まとめと今後の課題

参考文献

[1] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. OpenAI Gym. *arXiv e-prints*, p. arXiv:1606.01540, June 2016.

[2] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. 2013. cite arxiv:1312.5602Comment: NIPS Deep Learning Workshop 2013.