

# 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が注目を集めており、実世界の問題だけでなくゲームにも盛んに応用されている。本研究ではゲームバランスといった抽象的な概念を深層強化学習や進化型計算といった工学的手法によって調整する手法を提案し、独自のトレーディングカードゲーム (TCG) 環境において数値実験し、有効性を定量的に検証した。

### 2. 要素技術

#### 2.1. Deep Q Network

Deep Q Network (DQN) [1] 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。ある状態における行動ごとの Q 値を推定することで、状態が連続値といった Q 学習の適用が不可能な環境においても学習可能としている。Experience Replay などの工夫により安定した学習を可能としている。

#### 2.2. Genetic Algorithm

Genetic Algorithm (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応度を持ち、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応度を持つ個体を見つける。しかし、探索序盤で同一の個体で占められる初期収束という現象が起こり解の多様性が失われる場合がある。

### 3. 提案手法

本研究では以下の 2 つの手法を提案する。

- DQN により、あるデッキにおける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 種類ずつ除いて対戦していくことでそのデッキ内の構築戦略下におけるカードパワーを定量的に測定する手法。
- DQN により得られた結果を基にデッキ内で調整するカードを限定し GA の解空間の次元を削減することで、調整するカードの枚数を最小限にしながら TCG 環境のゲームバランスを調整する手法。

### 4. 実験

#### 4.1. 実験 1

学習側のデッキに恣意的に強いカードを 1 種類、弱いカードを 1 種類ずつ入れ、TCG 環境において DQN を適用した。後攻プレイヤーの行動を学習し、学習後 10000 回対戦を実行した。そこから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 DQN が適用できているかどうか確認した。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序盤のエージェントと比較することで学習済みエージェントの行動を分析した。結果として、0.7182 とベースラインと比較して高い勝率を残すエージェントを作成した。また、エージェントの行動を分析すると盤面のカードで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おり、恣意的に入れた強いカードは学習序盤と比較して優先的に用いられるなどカードの強弱も学習していると考察した。

表 1: 各手法で得られた最も良好な個体の適応度

手法 \ 評価指標	変更量	勝率	調整枚数
単目的 GA	0.44933	<b>0.85146</b>	0.36788
多目的 GA	<b>0.66365</b>	0.79097	0.42035
提案手法	0.53259	0.80783	<b>0.44933</b>

#### 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 種類ずつ除いて 10000 回ゲームを実行して先攻側の勝率を記録した。また比較対象として、ルールベースで作成した AI について同様に実験をし結果を比較した。結果として、デッキ内で最も強いと判断できるカードは同じ種類のカードであった一方で最も弱いと判断されたカードは異なる種類のカードであった。これは戦略の違いや学習済エージェントがカードの強弱を学習していることから、学習済エージェント同士の対戦において恣意的に弱く設定したカードは登場回数が少なく勝率計算に及ばず影響が少なくなったためと考察した。このことから深層強化学習を用いて未知のデッキにおける最適な戦略を構築しその戦略下におけるカードパワーを測定することで、より人間のプレイに近いシミュレーションデータを得ることができると考えられる。

#### 4.3. 実験 3

TCG 環境に実験 1 から用いてきたデッキを追加するといった問題を設定し、追加するデッキ内のカードのパラメータを環境内のデッキ間の勝率がそれぞれ 50 % に近づくように GA を用いて調整した。デッキ間の勝率の最適化において、実験 2 のデータから調整するカードの優先順位を付けて提案手法を適用した。また、比較手法として関連研究 [2] で用いられた勝率を最適化する単目的 GA、勝率とパラメータの変更量を最適化する多目的 GA を適用した。表 1 に結果を示す。各解において、勝率、パラメータの変更量、調整するカードの枚数について適応度を定義した。適応度が大きいほどその項目について優れた解である。各手法において各評価指標で優越する解を得ており、提案手法では調整するカード枚数に関して他の手法より優越した解を得ることができており提案手法の有効性が示された。

### 5. まとめと今後の課題

本研究では、深層強化学習を用いた TCG 環境の最適化手法を提案し、数値実験により有用性を示した。今後の課題として、DQN とは異なる深層強化学習手法の適用、GA 以外の離散最適化手法の適用、各アルゴリズムにおける最適なハイパーパラメータの発見などが挙げられる。

### 参考文献

- [1] 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.
- [2] Fernando de Mesentier Silva, Rodrigo Canaan, Scott Lee, Matthew C. Fontaine, Julian Togelius, and Amy K. Hoover. Evolving the hearthstone meta. *CoRR*, Vol. abs/1907.01623, , 2019.