

トレーディングカードゲームにおける 遺伝的アルゴリズムを用いた個性あるエージェントの作成と ニューラルネットワークによる模倣

山田 豊大^{†1} 阿原 一志^{†1}

概要: 不完全情報ゲームの 1 種であるトレーディングカードゲーム(TCG)は、デッキ(ゲームに使用するカードセット)を各プレイヤーが一定のルール化で自由に作成できるなど、囲碁や将棋、その他ボードゲームにはないゲーム要素が特徴である。したがって、強い、または人間らしいエージェントの作成等の研究は人間の意思決定過程の解明等に有意義であると考えられる。本稿では、『エージェントの個性に合わせたデッキを構築することが可能である』という仮説の実証のため、単純な TCG のプラットフォームをオリジナルに実装した。その上で遺伝的アルゴリズムによって個性のあるエージェント作成し、異なる個性のデッキを扱った際の挙動について観察した。さらに、そのエージェントの盤面評価をさらに良いものにするためにニューラルネットワークを用いたエージェントについても考察した。

キーワード: 不完全情報ゲーム トレーディングカードゲーム 遺伝的アルゴリズム ニューラルネットワーク 強化学習

Developing Agents by Genetic Algorithm and Imitation by Neural Network in Trading Card Game

ATSUHIRO YAMADA^{†1} KAZUSHI AHARA^{†1}

Abstract: Trading Card Game (TCG), one of an incomplete information game, has characteristic features. For example, any player can choose cards they use in their deck. Therefore, it will be significant for human's process of decision making to study how we obtain a smart or believable agent. In this paper, we developed characteristic agents by genetic algorithm, then we observed whether the agent can use the deck which has other characteristics well. After that, we make the agent's value function better, we considered an agent with Neural Network.

Keywords: Incomplete information game, Trading Card Game, Genetic Algorithm Neural Network, Reinforcement Learning

1. はじめに

トレーディングカードゲーム(TCG)は、「Magic: The Gathering」などが例として挙げられる、囲碁や将棋のようにゲームはターン制で進むが、手札やデッキ(各プレイヤーが使用するカード群)が非公開な不完全情報ゲームの 1 種である。近年ではコレクタブルカードゲーム(CCG)として「Hearth Stone」などがスマートフォンのアプリで遊べる。デッキを各プレイヤーが一定のプールとルール下で自由に構築ができ、カードの引きに行動が左右されるため、デッキに使用するカードの枚数配分が重要であり、大きな特徴である。そのため、デッキ構築の最適化とプレイの最適化、両方の観点からの研究が必要である。筆者らは、TCGにおいて高い性能を持つ AI や人間らしい AI を開発することは複雑な環境における人間の意思決定の理解に貢献すると考えている。

先行研究として[1]では Magic: The Gathering において、エージェントの動きを MCTS と Determinization を用いて最適化させており、[2]では遺伝的アルゴリズムを用いてデッキを強化している。

しかしながら、これまでの研究ではエージェントの行動

とデッキの作成、どちらか片方に軸を置いているものが多い。そのため筆者らは[3]で双方の相関を研究することを提案したが、一方のみを最適化することの問題点を指摘することとなり、利点については言及しなかった。

双方の相関を研究することで、デッキの作成といった観点では「よりエージェントに合ったデッキの作成」が、エージェントの行動という観点では「自分が使っているデッキの強い動きの学習」が可能になると考えられ、筆者らは「デッキを作成する際、そのエージェントの個性が反映される」、「エージェントの個性にあったデッキを使うことでより上手にプレイできる」という仮説を立てた。本稿ではこれらの仮説を検証するために単純な TCG のプラットフォームを作成し、そのプラットフォーム上で遺伝的アルゴリズムを用いて特定の個性を持つデッキ群の扱いに長けたエージェントを作成した。そして、一般の異なる個性を持つデッキを使用した際に、その個性を活かせるのかという検証を行った。さらに、盤面の評価をより適切にするためにニューラルネットワークを用いて盤面を評価するエージェントの作成の前段階として、遺伝的アルゴリズムを用いたエージェントの盤面評価を回帰により学習するエージェ

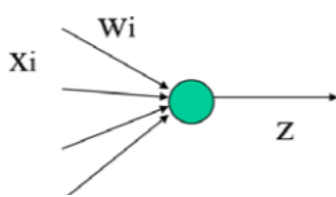
ントの作成を行った。

2. 関連研究

2.1 遺伝的アルゴリズム

遺伝的アルゴリズム[4]は John Holland によって提唱されたアルゴリズムで、選択と交叉、突然変異により評価関数を最大にする最適解にたどり着くことを目指すものである。本研究では盤面の評価関数の係数を最適にすることを目標とした。なお、後述のように本研究では交叉を用いずに要素のある範囲でランダムに変化させた。

2.2 多層ニューラルネットワーク



ニューラルネットワークは、生物の神経系における情報処理をコンピュータ上で模倣したものである。

ニューロンにおける入力と出力を組み合わせることで複雑な処理を実行する。前のニューロンから送られた入力 x_i は w_i の重みをもってニューロンに到達し、 z として出力される。 z は

$$z = f(\sum_{i=0}^N x_i w_i - \theta)$$

であらわされる。なお、 θ は閾値、 f は活性化関数である。詳しくは[5]を参照。

3. 実験内容

3.1 実験環境

実装は Python による。プラットフォームの作成に当たっては、TCG の複雑性を失わない程度に単純にするため一般のゲームにみられるようなヒーロー、サイドは採用せず、メインフェイズと戦闘フェイズを分けることはしなかった。以上を踏まえてヒーローのない Hearth Stone に準拠したルールに設定した。本研究で用いたコードと詳しいゲームのルールは[6]を参照。

3.2 実験内容

3.2.1 遺伝的アルゴリズムによる個性あるエージェントの作成(実験 A)

実験 A では、攻撃的と思われるデッキ群を準備し、これを学習して攻撃的なエージェントが作成できるかを実験した。このエージェントは、現在の状況から 5 行動後、または自分のターンの最後までを全探索し、その時の盤面を評価する。その際、盤面の要素をベクトル化して(対戦相手のライフ、自分の盤面にいるクリーチャーの情報など)、このベクトルとエージェントが持つ重みベクトルの内積を評価としている。この重みベクトルを、遺伝的アルゴリズムを用いて最適化した。なお、重みベクトルは単位ベクトルに限ることにより交叉は用いないこととし、それぞれの要素を[0,0.1]の範囲で調整することにした。最適化の手順を以

下の I から V に示す。

- I. 重みベクトルの候補をランダムに 32 個生成する。価値ベクトルの履歴集合にランダムに生成した単位ベクトル 1 つを追加する。
- II. 重みベクトルの候補 1 つにつき 10 回価値ベクトルの履歴集合からランダムに選んだ重みベクトルを持つエージェントと対戦をし、勝利数を記録する。
- III. 勝利数が最も多いベクトルをその世代で採用する重みベクトルとし、このベクトルとその次に勝利数が多いベクトルを履歴集合に追加する。
- IV. 追加したベクトル 2 個とそこから要素を調整したベクトル 28 個、ランダムに生成したベクトル 2 個計 32 個を候補とする。
- V. II から IV を繰り返す。

序盤から攻撃をするのに適しているようなデッキ群を作成し、このデッキ群に関してエージェントを強化した。その後、序盤は守備的にプレイするような個性の異なるデッキ群を与え、強さやプレイスタイルを観察した。

3.2.2 ニューラルネットワークによる模倣(実験 B)

本実験は実験 A の結果を受けて行った。実験 A で作成したエージェントは盤面の評価関数が線形変換であるため、常に盤面の要素の重みが一定である。しかし、ターンの経過とともに重要になってくる要素もあると考えられる。したがって、「隠れ層を追加し、より盤面の評価を適切にできるエージェント」を目的とし、ニューラルネットワークを用いたエージェントの構築を考えた。その前段階として、まずは遺伝的アルゴリズムによるエージェントの行動を、ニューラルネットワークを用いて学習することにした。

そこで実験 B では、最終的に得られた重みベクトルの値を持つエージェントの行動を、ニューラルネットワークを用いて学習することを行った。まず、ランダムに行動するエージェントまたは自分自身との対戦を 300 回行い、盤面の評価値と盤面ベクトルを記録した。盤面ベクトルを入力、評価値を出力とする $128 \times 40 \times 20 \times 1$ のニューラルネットワークを用いて実験 A のエージェントの盤面評価を学習した。学習にあたっては対戦で得られたデータから重複した要素を省いた約 35000 件のデータをシャッフルし、1000 件ずつのデータに分割した後、5 エポック分学習させながらヒューリスティクスのエージェントと 50 回対戦を行いその成績を記録した。盤面として以下の要素を記録した。

表 1 教師データの記録要素

Table 1 Factors of Training data

要素	次元
お互いのライフ	12×2(ライフの値を 0~11の値に抑えている)
自分の手札にいるカードの種類	16×2(カードの種類×デッキに入る枚数)
自分の手札の枚数	5(0 から 4 に抑えている)
自分の盤面	18(クリーチャーのパワーやタフネスをある閾値で分けている)
自分の盤面にいるクリーチャーの合計パワー	10 (0 から 9 に抑えている)
対戦相手の盤面	18
対戦相手の盤面にいるクリーチャーの合計パワー	10
対戦相手の手札の枚数	5
自分が現在使えるマナ	6
合計	128

4. 結果と考察

4.1 実験 A

強化を図ったエージェントの強さについて述べる。強さの指標は、筆者が設定した重みベクトルを持つヒューリスティクスエージェントと 50 回対戦させた時の勝率を用いた。なお、このエージェントは完全にランダムに行動するエージェントに対し 50 戦中 50 勝することができた。筆者(人間)との対戦では筆者が勝ち越したが、不自然な手を打つことは少ないという印象を受けた。結果を図 2 に示す。

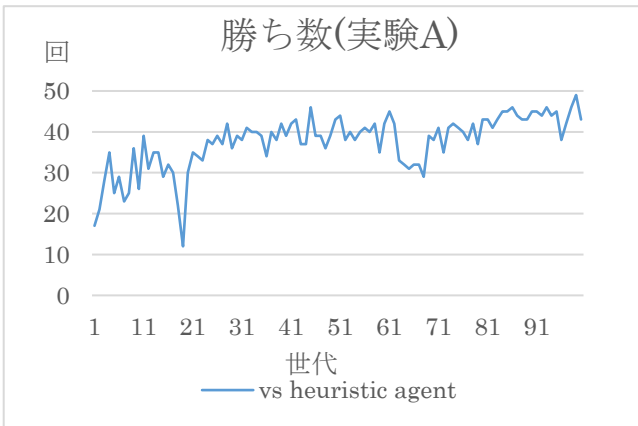


図 2 世代における勝利回数の推移

Figure 2 Transition of number of win at the generation

50 世代を超えるあたりでヒューリスティクスエージェントに対して最低でも 6 割以上、80 世代を超えるあたりではおおよそ 8 割の勝率を上げることができるようになり、十分に強化されたと考えられる。このエージェントに個性

の異なるデッキを扱わせたと、100 戦中 22 勝しかできなかった。行動を観察すると、序盤は守りのプレイをしなければならぬデッキであるにも関わらず、強化させた際に用いたデッキが攻撃的なデッキであるため、序盤から無理に攻撃を仕掛けている展開が多かった。負けた試合の 7 割は守りが薄くなったところを攻め込まれ、残りは純粋に攻撃性能に劣っているため攻めきれずデッキがなくなって敗北した。

以上のことから次のような考察を行った。TCG においてはデッキに個性を付与して学習すると、プレイヤーにも個性が現れ、個性の異なるデッキには対応できなくなる。したがって、この個性あるプレイヤーがデッキを作成した際、自分に合ったデッキを作成しようとする、攻撃的なデッキを選別できるようになると考えられる。

4.2 実験 B

このエージェントの強さに関して、ヒューリスティクスのエージェントとの対戦結果を図 3 に示す。

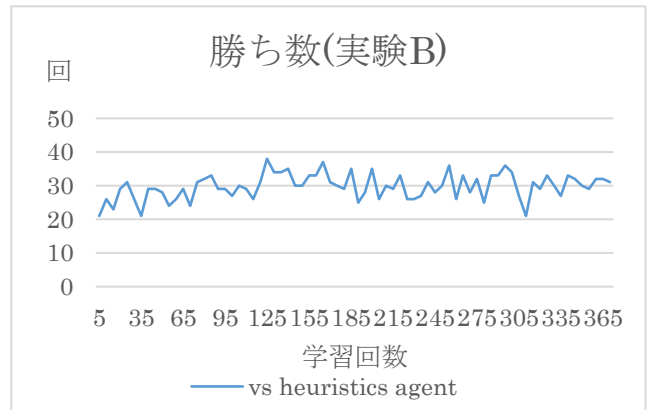


図 3 学習回数における勝利回数の推移

Figure 3 Transition of number of win at the number of learning

学習回数が 65 回までは勝利数が 30 に届かないことが多いが、学習回数が 150 回を超えるあたりで勝率が 6 割近くに収束している。元のエージェントがヒューリスティクスのエージェントに対して 8 割の勝率を挙げていることを踏まえると完全に学習しているとは言えないが、勝ち越しているところをみると学習している様子が見て取れる。したがって、ニューラルネットワークを用いて盤面を評価するエージェントの作成も可能ではないかと考えられる。勝率が 8 割に届いていない原因としては、ニューラルネットワークの規模に対してデータ数が少ないために学習できなかった盤面があるのではないかと考えられる。

しかしながら、本実験においては実験 A で作成したエージェントを学習するために実験 A で作成したエージェントが用いている盤面要素をニューラルネットワークの入力情報としたが、自分の手札の枚数は自分の手札にいるクリーチャーの情報から算出できるなど、盤面ベクトルの要素の

取り方がやや恣意的であることは注意すべき点である。一般の場合には盤面の情報のみから評価をすることが肝要であると考えられるため、恣意的な情報を抜いた際の挙動についても観察する必要がある。また、本実験で作成したエージェントはあくまで盤面情報を線形変換して評価をするエージェントを学習したものであるため、隠れ層を追加した意味が薄い。今後はさらに良いエージェントの模倣、教師なしでも強くなるような方法を探したい。

5. まとめと今後の展望

本稿では、個性あるエージェントがデッキを作るとその個性がデッキに現れる、エージェントの個性と合ったデッキを用いることでより強いエージェントになるという仮説を立てた。その検証のため、個性をつけたエージェントに異なる個性のデッキを扱わせる実験と、盤面評価をより詳細に行うエージェントの作成を目指してニューラルネットワークを用いて遺伝的アルゴリズムを用いたエージェントの盤面評価を回帰により学習するエージェントの作成を行った。

本研究の課題として、実験 A でエージェントに与えたデッキが 2 つとも筆者の主観で作成している点があげられる。実験 A で守備的なデッキを使った際にエージェントが勝利できなかった原因が、「2 つのデッキの間に強さの差があり守備的なデッキが弱い」という点である可能性があるためである。この点は、「ルールベースに対して同じような勝率を持つデッキ」など予め同じくらいの強さであると考えられるデッキを使って検証することが必要だと考えられる。また、実験 B ではあくまで盤面要素をヒューリスティック的に入力データにしているため、盤面要素をそのまま入力にした際の挙動についての観察や、教師あり学習の枠にとどまっているため教師なしで強化学習する手法の検討が必要であると考えられる。今後は、本稿の課題の解消、実際にエージェントにデッキを作成させる方法やその評価方法の確立、他の研究者が使いやすくなるようなプラットフォームのリファクタリングを行いたい。

6. 参考文献

- [1] Peter I. Cowling [ほか]. (2012). “Ensemble Determinization in Monte Carlo Tree Search for the Imperfect Information Card Game Magic: The Gathering”. *IEEE Transactions on Computational Intelligence and AI in Games* (241-257),4(4)
- [2] Pablo García-Sánchez [ほか]. (2016). “Evolutionary deckbuilding in hearthstone”. *IEEE Conference on Computational Intelligence and Games* doi: 10.1109/CIG.2016.7860426
- [3] 山田豊大, 阿原一志.(2018) トレーディングカードゲームにおけるデッキ作成とエージェント構築を目標

としたニューラルネットワークを用いた学習モデルの検討,(2018) GPW-18(128-132)

- [4] Holland J.H. (1984) *Genetic Algorithms and Adaptation*. In: Selfridge O.G., Rissland E.L., Arbib M.A. (eds) *Adaptive Control of Ill-Defined Systems*. NATO Conference Series (II Systems Science), vol 16. Springer, Boston, MA
- [5] 斉藤康毅.(2016) 「ゼロから作る Deep Learning - Python で学ぶディープラーニングの理論と実装」
- [6] github. (https://github.com/Yamada-maya/Flex_cardgame-Arena, 2019/07/17 現在).