

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 ニューラルネットを用いた学習モデルの検討

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)をプレイヤーが選択できるなど、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り、強い、又は人間らしいエージェントの作成等の研究には意義があると考えられる。本稿は、単純化されたTCGで、デッキ作成、またエージェントの構築に関する知見を得ることを目標とし、その第一段階としてランダムに与えられたデッキに対してニューラルネットワークを用いた強化学習の有効性について考察した。

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 ニューラルネットワーク 強化学習

Consideration of Learning Model with Neural Network to Build Decks and Develop Agents 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 to study how we obtain a smart or a believable agent. In this paper, we set goals to gain the knowledge of building decks and developing agent in simplified TCG. As the first step, we considered the effectivity of Neural Network by applying that to the games where both players are given random decks.

Keywords: Incomplete information game, Trading Card Game, 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」や「遊☆戯☆王オフィシャルカードゲーム」などが例として挙げられる不完全情報ゲームの1種である。近年では「Shadow Verse」や「Hearth Stone」といったスマートフォンのアプリとしても遊べるようになっており、その認知度は高まっている。TCGは囲碁や将棋のようにターン制で進むが、使用カードを各プレイヤーが選べ(使用するカードの束をデッキと呼ぶ)、デッキからランダムに引くカードの種類、いわゆる「引き」によって、戦局や取れる行動が左右されるという点が異なり大きな特徴である。

TCGは主に2つのステップ、「デッキを作成するステップ」と「実際に対戦するステップ」によって構成される。デッキを作成することは戦略を決めることに相当し、この作業がうまくできることは重要である。一方で、実際の対戦で適切な行動をとることはその戦略を十全に生かすことであり、こちらもまた重要である。そのため、囲碁や将棋といった完全情報ゲームや、ポーカーといった不完全情報ゲームで重要視される「エージェントの動きを最適化させる」「どのようなカードでデッキを作るのか」という複眼的な視点でゲームを考察する必要があると考えられる。先行研

究として[1]では Magic; The Gathering において、MCTS と Determinization を用いてエージェントの動きの最適化をa行っており、[2]では遺伝的アルゴリズムを用いてデッキを強化することをやっている。

ところが、デッキ作成とエージェントの行動双方に目を向けたゲーム全体としての学習モデルは未だに確立されていない。一見、これらは独立した作業でそれぞれ別個に確立すればよいと考えられる。すなわち、上記の研究をさらに深化させてその結果をそれぞれ適用すれば最適なエージェントが完成する、という考えである。しかし、どちらか一方だけでは以下に述べるような懸念事項が出てきてしまう。まず、デッキ作成を深化するにしても、デッキを作っただけではその強さは測りづらい。そのため、そのデッキを使用するエージェントが必要となる。その際、“そのエージェントにとって強いデッキ”のみが完成してしまう懸念がある。一方、エージェント構築を深化させるにしても、特定デッキのみ扱えるエージェントができてしまう懸念が考えられる。

なお、[3]のように、現在スマートフォンアプリで用いられているエージェントは恣意的な思考ルーチンに基づいており、TCGの意思決定に関する知見の一般論は得られない可能性がある。

^{†1} 明治大学 総合数理学部 先端メディアサイエンス学科
Department of Frontier Media Science, School of Interdisciplinary Mathematical

Sciences, Meiji University

デッキ作成とエージェントの行動双方に目を向けることにより、エージェントはデッキに対してどのような行動をとると強力であるかが学習可能であり、またデッキを作成するにあたって、エージェントの行動と戦績を踏まえたデッキを作成することが可能になると考えられる。これはいわゆるメタゲームと呼ばれる考え方であり、特に TCG の大きな大会で勝利するために必要なデッキの構築手法である。

以上のことから、TCG の強い、または一部のカードのみでデッキを作り対戦する、というような何らかの条件に基づいたエージェント構築は研究内容として意義深いものと考えられる。本稿では、最終的なゴールとして TCG のエージェント構築と(前述のような何らかの制約を課したうえでの)効果的なデッキ作成を見据えつつ、その第一段階としてプラットフォームの作成とデッキ確定後のエージェントの学習手法を提案し、考察を行った。

2. 関連研究

2.1 多層ニューラルネットワーク

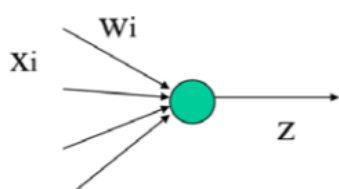


図 1 多層ニューラルネットワーク
Figure 1 Multilayer Neural Network

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 は式(1)であらわされる。

$$z = f(\sum_{i=0}^N x_i w_i - \theta) \dots (1)$$

θ は閾値、 f は活性化関数である。詳しくは[4]に記述されている。

3. 実験内容

実装は python による。本研究では以下のルールのゲームを実装した。用語については世界的にもプレイヤーの多い Magic: The Gathering の用語を用いたが、プレイヤーの行動をできるだけシンプルにすることを考えてルールの内容については shadow verse や hearth stone のものを参考に構築した。さらに詳しいルールは 6 章に譲る。

- 1). プレイヤーは 2 人。個性の異なる 15 種類、各 2 枚ずつ計 30 枚のカードの中から 20 枚以上カードを選び、それぞれのデッキとする。
- 2). 先攻を決める。各プレイヤーはライフ(life)とマナ mana という 2 つのステータスを持つ。カードはそれぞれコスト(cost)、攻撃力(power)、体力(toughness)を

持つ。一部のカードはスキル(skill)として特殊な効果を持つ。

- 3). 先攻は 2 枚、後攻は 3 枚、それぞれよく混ぜられ、順番がランダムになっている自分のデッキからカードを引き、それぞれの手札とする。
先攻はターンプレイヤーとなる。
- 4). ターンプレイヤーはすべての場のカードをアンタップしてデッキからカードを 1 枚引き、マナが 5 でないならマナを 1 増やしてマナを回復する。この時、デッキからカードが引けなかったプレイヤーはライフが 0 になる。その後マナを消費してカードをプレイする。プレイしたカードはタップ状態で場に出される。
- 5). タップ状態でないカードはタップすることで相手か相手の場にあるカードを攻撃できる。この時、カードの体力が 0 以下になったカードは墓地に送られる。
- 6). ターンプレイヤーでない側にターンプレイヤーが移る。上記 4)~6) をどちらかのライフが 0 になるまで繰り返す。

プレイヤーが思考し、それを結果に移せるフェーズは 4). と 5). であるが、本研究では 4). の有無によるゲームや学習の難易度を測るためにカードのプレイをルールベースにしてプレイヤーの行動を学習する手法と、カードのプレイ・プレイヤーの行動双方を学習する手法を試した。

まず、ランダムに行動するプレイヤー同士で 20,000 回対戦させ、その対戦を記録した。記録した内容は、それぞれ以下の通りである。

表 1 教師データの記録要素

Table 1 Factors of Training data

記録要素	次元数
自分の手札	15×2
自分の場	15×2×2
対戦相手の場	15×2
自分の山札の残り	15×2
お互いの墓地	15×2×2
お互いのライフ	20×2
合計次元数	250

記録要素	次元数
プレイしたカードの種類	15
カードをプレイしたかどうか	1
合計次元数	16

記録要素	次元数
行動したカードの種類と攻撃先	15×2
カードが行動したかどうか	1
合計次元数	31

カードの攻撃先の記録要素は試験的に「対戦相手か場に出ているカードか」という大雑把な分け方をした。本来であればどのカードに攻撃をしたか、という情報まで記録すべきであるが、まずは少しでも学習ができるかどうかを計測することにした。

その後、勝利したプレイヤーの行動を教師データとして3層のニューラルネットを用いて学習させる。カードのプレイを学習するニューラルネットワークは(250×250×16)、プレイされたカードの行動を学習するニューラルネットワークは(250×250×31)にした。その際、本研究においてプレイヤーの行動は状況という入力に対する出力であると考えた。そのため、ニューラルネットワークでの教師データにする際に行動の種類が均等になるようにデータを加工した。この学習手法では自分のデッキのデータを思考に組み込んでいるためデッキに応じた行動が可能であると考えられる。

そして100回学習するごとにランダムに行動するエージェントとルールベースのエージェントとの100回対戦を1セットとしてそれを10セット行い、勝率を記録する、ということをして70ループ行った。それが終了したとき、追加で25戦行いその戦いの経過をファイルにして記録、行動の観察を行った。

学習を行ったエージェントは対戦において、行動の度に現在の状況を入力としてすべての行動に対するsoftmax値を計算し、最もsoftmax値が高かった行動を出力する。その行動を行うことができなかつた場合、行動をパスする。

さて、どのカードをどのタイミングで使用するか、ということはカードゲームの中でも重要な要素である。仮に手札が3枚、マナが無限に使用できるとするとどのカードを使用するかという組み合わせは 2^3 通りで済むものの、Magic; The Gatheringのようなゲームにおいては場合によっては8枚以上の手札をため込むことも考えられる。そのことを考えると、エージェントにカードのプレイを学習させることはカードの行動を学習させることと同程度の労力や時間を要求すると考えられる。

本稿で設計したゲームは単純であり、ルールベースでカードのプレイは記述できる。そのため、カードの行動の最適化の研究題材としては適していると考えられる。本来であれば本稿はカードの行動の最適化に絞って考察すべきだったかもしれないが、一方で、Magic; The Gatheringのよ

うにカード種が多種多様で使用する順番やタイミングが重要なゲームも存在する。今後そのようなゲームを研究題材にするにあたり、カードのプレイに関する学習も行った。

4. 結果と考察

本研究では、対戦相手としてランダムに行動するエージェントとできるだけ早く相手の体力を削ろうとするルールベースのエージェントの2つを用意した。ルールベースのエージェントの行動ルーチンは以下の通りで、これはMagic; The Gathering や shadow verse においてアグロと呼ばれる戦術を本研究用にアレンジしたものになる。

- 1). カードをプレイする際は、コストの大きいカードからマナがなくなるまで使用する。
- 2). カードを行動させる際、攻撃の値が大きいカードから順に行動させ、
 - (ア) そのカードが倒しきることのできるカードが存在する場合、そのカードを攻撃する。
 - (イ) そうでなければ、相手のプレイヤーを攻撃する。

以下に記したグラフはそれぞれの対戦エージェントとの学習回数に応じた勝数の平均を示している。

4.1 ランダムに行動するエージェントとの対戦

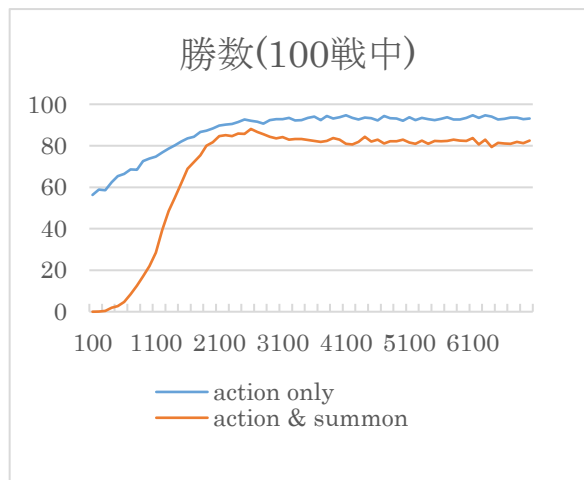


Figure 2 Results of competing with Random agent

手札のプレイ(summon)・行動(action)双方を学習したとき、また行動のみを学習したとき両方において、約4000回学習した時点で近似曲線の傾きが0に近くなっており、学習が完了したことがわかる。その時の勝数の平均は前者で8割、後者で9割となっており、学習の効果があつたと考えられる。さらに以降の勝数の標準偏差は5より小さく、有意に勝ち越していると考えられる。前者の勝数のグラフを見ると、初期の勝数がほぼ0であるが、学習を重ねるにつれて高い勝率となっており、学習の効果が顕著に表れている。

行動を観察してみたところ、カードをプレイし場にカード

を出す、という点では学習初期によく見られた「まったくカードをプレイせずに手札にため込む」という行動は見られず、1 ターンに複数のカードをプレイすることもあるなど学習の効果が得られていることが確認できた。一方で、カードの行動についてはやや特殊な傾向が見られた。カードをプレイした後、行動できるカードは相手を直接狙う前に先に相手の場にあるカードを狙っていた。すなわち、必ずしも適切な順序でないにせよ、(もっと被害が少ない攻撃の順番がある)相手の場にある自分を狙えるカードを丁寧に処理する傾向が見られた。そして、自分の場が盤石になったときようやく相手を攻撃していた。その際確実に行動できるカードを行動させているわけではなく、326 ターン中 137 ターンは行動できるカードを攻撃させないままターンを終了することもあり、学習したもののまだまだ行動の最適化がなされていないと感じられた。この傾向は 25 戦すべてにおいて観察され、例外は相手を直接攻撃するとその時点で試合に勝利する、という場面のみであった。

4.2 ルールベースのエージェントとの対戦

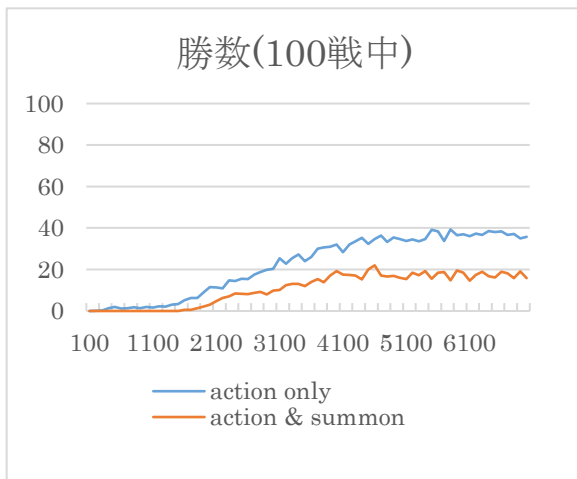


図 3 ルールベースエージェントとの対戦結果

Figure 3 Results of competing with Rule based player

勝数が最初は 0 だったのが徐々に上昇し、学習はしているものと考えられる。しかしながら、肝心の勝数は 5 割を切っており、強力なエージェントを作る、という目的は達成できなかった。それでは、本稿のアプローチが無効だったのか、を考察していく。

行動を観察してみたところ、カードをプレイして相手を直接攻撃する、その前に相手の場に出ているカードを攻撃し、そうしているうちに結果的にデッキからカードを引けないことによる勝利につながっていることが多い印象を受けた。具体的には行動観察用の対戦譜のうち行動のみを学習させたエージェントは 7 勝を挙げているが、そのうちのすべてにおいてデッキ切れで勝利している。また、そのうちの 5 回は先攻での勝利である。先行は先にカードを展開できるので、相手の場のカードの処理が後攻の時よりも容易だったと考えられる。敗北した試合は相手が積極的にラ

イフを攻撃してくるのでそれを捌ききれず負けている、という展開となっていた。ゲーム終了時、相手のライフが 20 のままであることから、相手の場のカードを処理することで手一杯となり、相手を直接攻撃する余裕がなかったということが見て取れる。一見すると単純に負けているように見えるが、攻撃を捌き切れていない要因の一つにカードが相手の場に出ているカードを攻撃するときに「相手の場のカードのどれかを攻撃する」というようにターゲットを大雑把にしか決めていない点がある。そのため、この点や相手を直接攻撃しない点を改善することでさらに勝率を上昇させられるのではないかと考えている。そのため、本稿のアプローチが完全に無効、とは言い切れない。

5. まとめと今後の展望

本稿ではランダムに与えられたデッキに対して、ニューラルネットワークを用いて学習を進めたエージェントがルールベースのエージェントに対してどのような勝率を上げるのか、そしてその理由を考察した。また、カードのプレイを学習することの重要性についても考察した。

今後はカードの行動、例えば攻撃する順番や攻撃先、の更なる最適化や、本稿で触れることのできなかったデッキ構築についてのアプローチの提案、さらにカード種が増えて入力次元が増えた際にも本稿のアプローチが有効なのかについても調査したい。

そのためのアプローチとして、教師データの整備や 2 枚のカードの組み合わせによる評価のようなものを考えていきたい。

6. ルールの詳細

6.1 カードの一覧

カード ID	コスト	攻撃	体力	スキル
0	0	1	1	
1	1	2	1	
2	2	3	2	
3	3	4	3	
4	1	1	1	速攻
5	2	2	1	速攻
6	3	3	1	速攻
7	1	2	1	治癒
8	2	1	4	治癒
9	1	1	1	蛮勇
10	2	1	3	蛮勇
11	3	2	3	蛮勇
12	1	0	3	強化
13	2	0	4	強化
14	3	1	4	強化

6.2 スキル一覧

6.2.1 速攻

このカードをプレイした時、このカードはアンタップ状態で場に出る。

6.2.2 治癒

自分のターン終了時、このカードの体力は1大きくなる。

6.2.3 蛮勇

自分のターンの間だけ、このカードの攻撃は2大きくなる。

6.2.4 強化

自分のターン終了時、このカードの攻撃は1大きくなる。

7. 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] Cygames Engineer’s Blog.
(<http://tech.cygames.co.jp/archives/2853/> 2018/07/27 現在)
- [4] 齊藤康毅. (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」