

# 対戦型カードゲーム AI におけるデータ拡張の研究

M2022SE001 古川湧也

指導教員：沢田篤史

## 1 はじめに

近年、対戦ゲームの分野において AI の研究が注目を集めている。対戦 AI は学習を重ねることにより熟練のプレイヤーの行動と同じ振る舞いを再現することができ、2015 年に囲碁の分野において囲碁対戦 AI である AlphaGo が棋士の名人に勝利している [1]。他にもユーザに合わせた難易度での対戦機能を実現する対戦 AI の研究も進められている [2]。

対戦ゲームの対戦 AI の学習に用いる訓練データを人手で収集するには多大な労力を要する。カードゲームではカードごとに異なるステータスが設定されているので、考え得る棋譜の組み合わせは膨大な数になる。加えて棋譜を機械学習の入力データとして扱うにはデータの順番、すなわち時間依存性を考慮する必要がある。さらに対戦 AI に多種多様な戦略を学習させるには訓練データとなる棋譜は戦略面を考慮した内容のものを用いる必要がある。

本研究ではカードゲームに着目し、カードゲーム対戦 AI の訓練に有効なデータを生成する仕組みを提案することを目的とする。対戦 AI に使用可能な棋譜データを訓練データとして機械学習の生成モデルを用いてデータ拡張を行う。

本研究の目的を達成するのに必要な技術課題を次に示す。

- GAN (Generative Adversarial Network) におけるカードゲームの棋譜データの適切なエンコードを明らかにする。
- 少量のデータから時系列を考慮したデータを効果的に生成する方法を検討する。
- エンコード手法および生成モデルを評価する。

機械学習では学習対象のデータの特徴を表すベクトルを入力とする。カードゲームの棋譜のベクトル化の方法を提案することで機械学習におけるカードゲームの棋譜データの適切なエンコードを明らかにする。GAN によるデータ拡張によって少量のデータから時系列を考慮したデータを大量に生成する。カードゲームの棋譜をゲームの進行と共に推移する離散型時系列データとみなし、それを GAN の訓練データとすることで大量の棋譜データを生成する。Lantao ら [3] の SeqGAN を参考に、GAN の学習に強化学習の枠組みを用いることで離散型データに対応した生成モデルを設計する。さらに生成された棋譜データおよび設計した離散時系列 GAN を評価指標を定義し、適切なベクトル化手法と離散時系列 GAN の性能を考察する。

本研究の技術課題に対するアプローチとして GAN を用いたデータ拡張を行う棋譜生成システムを実装する。カードゲームの棋譜をゲームの進行と共に推移する離散型時系列データとみなし、それを棋譜生成システムの訓練データとすることで棋譜データを生成する。カードゲーム対

戦 AI およびそれを動作させるカードゲームプログラムを実装し、棋譜生成システムを評価する。棋譜生成システムが生成した棋譜のみをカードゲーム対戦 AI の学習に使用し、カードゲーム対戦 AI の挙動によって棋譜生成システムが生成した棋譜の妥当性を評価する。

結果として棋譜の生成、対戦を行うことができたが望むような結果は得られなかった。その原因としてデータ形態の複雑性、ルールの簡素さによる棋譜の偏りが考えられる。

本研究の成果によってカードゲーム対戦 AI の実現可能性が向上する。多彩な戦略を学習するための訓練データとなる棋譜データを大量に生成することで、カードゲーム対戦 AI の実現を支援することができる。

## 2 背景と関連研究

### 2.1 背景

対戦ゲーム AI はゲームの進行状況を記録した棋譜を用いて、そのゲームにおける最適な行動を学習する。対戦型カードゲームでは複数の種類のカード、カードの持つステータス、カードを配置する盤面の位置などの要素から考えられる盤面の組み合わせは膨大に存在する。そのような棋譜の収集を人手で行うには多大な労力を要する。加えて使用するデータセットはそのゲームのルールに則っていたり、熟練者の動きを記録した棋譜データを用いるのが望ましい。そうでないものを教師データとして対戦ゲーム AI の学習に使用した場合、その対戦 AI はゲームのルールにそぐわない挙動や勝利につながらない行動を起こす恐れがある。

### 2.2 関連研究

離散型時系列データを対象としたデータ拡張の関連研究として Lantao ら [3] の SeqGAN (Sequence Generative Adversarial Nets) がある。生成モデルによって離散型時系列データを生成する場合、次の 2 つの問題点が存在する。

- 学習に誤差逆伝搬を用いることができない
- 識別器は時系列データ全体しか評価できない

連続的なデータの GAN は、生成器の生成データを識別器に入力し分類誤差を計算、それを誤差逆伝搬により生成器のパラメータを調整する。しかし訓練データが離散的である場合データ間に連続的な特性は存在せず勾配が存在しない。よって誤差逆伝搬を用いた訓練が不可能である。識別器は訓練データもしくは生成器の出力データを入力とする。すなわち時系列データ全体のみを入力とするので、ある時間の時系列データがそれ以降の時間においてどのような影響を及ぼすのかを考慮した学習が不可能である。

SeqGAN では GAN に強化学習の枠組みを適用し離散型時系列データである自然言語を生成している。離散型

時系列データの特徴を学習するために生成器が出力した複数個の文章を識別器に入力し、その予測結果の平均から生成器のパラメータを調整する。

### 3 研究目的と技術課題

#### 3.1 研究目的

本研究の目的はカードゲーム対戦 AI の訓練に有効なデータを生成する仕組みの提案である。SeqGAN の手法を参考に、GAN の学習に強化学習の方法論を適用し離散型時系列データのデータ拡張を行う。ある時間の盤面と自身の手札の系列をカードゲームにおける棋譜とし、それをベクトル化することで本研究の提案手法のデータセットとする。カードゲーム対戦 AI に学習させ、実際に対戦を行う。カードゲーム対戦 AI の予測結果からエンコード手法と生成モデルを評価する。

#### 3.2 技術課題

カードゲーム対戦 AI の訓練に有効なデータを生成することの課題は次の 3 点である。

- GAN におけるカードゲームの棋譜データの適切なエンコードを明らかにする。
- 少量のデータから時系列を考慮したデータを大量に生成する方法を検討する。
- エンコード手法および生成モデルを評価する。

機械学習では対象のデータの特徴を学習するための教師データは数値化したベクトルで表現されている必要がある。カードゲームの棋譜データは数値化されておらずそのままの状態では機械学習に活用できない。よって機械学習の生成モデルである GAN に活用できるようなカードゲームの棋譜データの適切なエンコードを明らかにする必要がある。

離散的な時系列を考慮したデータの生成方法を検討することで離散型時系列データであるカードゲームの棋譜を生成する。カードゲームの棋譜を離散型時系列データとして扱い既存の方法論から少量のデータから時系列を考慮して大量にデータを生成する方法を検討する。

提案したエンコード手法及び生成モデルを評価することでカードゲーム対戦 AI の訓練に有効なデータを生成できているか確認する。カードゲームの棋譜をベクトル化するにあたってカードゲーム対戦 AI が意図した挙動を行うのに必要な情報を確かめる。また生成モデルの最適なハイパラメータおよび生成された棋譜の妥当性から生成モデルの性能を評価する。生成モデルの学習状況、対戦カードゲーム AI の挙動から生成モデルを評価する。

## 4 カードゲームの棋譜生成システム

### 4.1 概要

本研究では、技術課題に対するアプローチとして棋譜のデータ拡張を行う GAN の生成モデルを設計する。棋譜生成システムの GAN のネットワークを図 1 に示す。

図 1 の GAN ネットワークはベクトル化したカードゲームの棋譜データを教師データとする生成モデルである。棋譜生成システムは SeqGAN の手法を参考に設計した。生

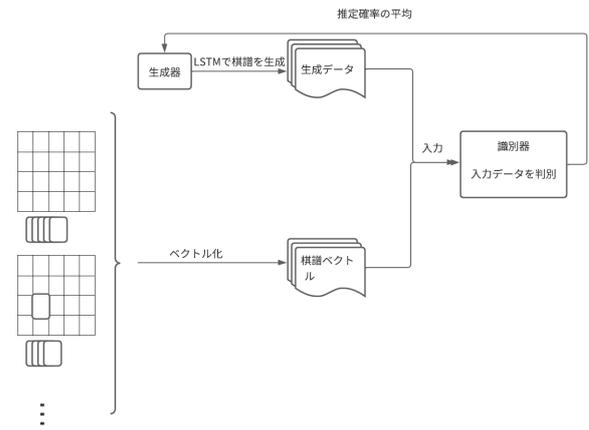


図 1 棋譜生成システム

成器は LSTM で離散型時系列データである棋譜データを生成する。識別器は 2 値の分類モデルで入力されたデータが訓練データか生成器が生成したデータか推定する。

### 4.2 カードゲームの棋譜のベクトル化

カードゲームの状況をニューラルネットワークに入力するのに最適な形にエンコードする。棋譜生成システムでは使用するカードすべてに別々の ID を割り当てる。割り当てた ID によって棋譜データ上でカードを識別する。プレイヤーは用意されたすべてのカードを使えるわけではなく手持ちのカードしか使用できない。加えてカード一枚一枚が異なるステータスを持って。なので場の状況を鑑みて出すカードを考慮しなければならない。このことからプレイヤー自身の手札と場の状況をベクトル化の対象とする。

### 4.3 生成器の学習

カードゲームの棋譜生成システムの生成器は LSTM を用いてカードゲームの棋譜を生成する。LSTM によって時系列データであるカードゲームの棋譜を生成する。カードゲームの棋譜の時間依存性を考慮してデータ生成を行うために一つの棋譜を異なる長さに分割し生成器に入力する。生成器の LSTM で入力棋譜の途中からゲーム終了までの棋譜を生成する。生成された棋譜を識別器に入力し本物の推定確率の平均から生成器のパラメータを更新する。この方法で離散型時系列データであるカードゲームの棋譜の特徴を時間依存性を含めて学習できる。

### 4.4 識別器の学習

カードゲームの棋譜生成システムの識別器は入力されたデータが訓練データか生成器が生成したデータか推定するように学習する。棋譜生成システムでは教師データは離散型時系列データであるので、生成器の学習の際には生成器が複数生成したデータを入力としその推定確率の平均から生成器のパラメータを調整する。

## 5 実験および実装

### 5.1 カードゲーム対戦 AI の実装

棋譜生成システムの生成する棋譜の妥当性を確認するためにカードゲーム対戦 AI を実装する。カードゲーム対戦 AI には PolicyNetwork[4] を用いる。PolicyNetwork は畳み込み層と全結合層によって現在の状況における最適な一手を予測する。カードゲーム対戦 AI への入力特徴量は自分の手札と場の情報を統合した配列を用いる。

### 5.2 カードゲームのルール

本研究のカードゲームのルールについては「遊 戯 王 オフィシャルカードゲーム」<sup>1</sup>のものを参考にそれを簡略化したものを定義した。棋譜生成システムで生成される棋譜はこのルールでの対戦の状況を記録したものである。

カードゲームのルールの概要を以下に記す。

1. プレイヤーは二人。それぞれ 4000 の LP (ライフポイント) という持ち点を持つ。
2. 各プレイヤーは 40 枚のカードを集めたデッキを持つ。
3. カードにはモンスター、魔法、罫の三種類がある。モンスターには名称、攻撃力、守備力、レベルのステータスを持つ。魔法、罫は名称、効果のステータスを持つ。
4. プレイヤーは自分のターンにデッキの一番上からカードを一枚引き手札に加える、カードをプレイし、自身の場のモンスターで攻撃を行い、ターンを終了する。
5. プレイヤーが自身の場のモンスターで攻撃を行う場合、攻撃対象となる相手モンスターとのステータスを比べその結果によってプレイヤーの LP が減少する。相手の場にモンスターが存在しないときモンスターの攻撃分の数値が相手の LP から引かれる。
6. プレイヤーがターンを終了すると相手のターンとなる。
7. どちらかのプレイヤーの LP が 0 になるまで互いのターンを繰り返し、LP が 0 になったプレイヤーは敗北となる。
8. デッキにカードが存在しない状況でカードを引いたプレイヤーは敗北となる。

### 5.3 カードゲーム対戦プログラム

カードゲーム対戦 AI の妥当性を評価するために上記のルールに従い対戦を行うプログラムを実装した。各プレイヤーがカードを場に出す度にフィールド配列を更新し、その状態を記録し棋譜とする。このプログラムではお互いに同じデッキを使用する。場は両プレイヤーそれぞれのモンスターゾーンが 5 つ、魔法・罫ゾーンが 5 つの 4 行 5 列の二次元配列で表現される。

### 5.4 カードゲーム対戦 AI の入力特徴量

カードゲーム対戦 AI への入力特徴量は自分の手札と場の情報を統合した配列を用いる。手札と場のベクトル化

を図 2, 3 に示す。

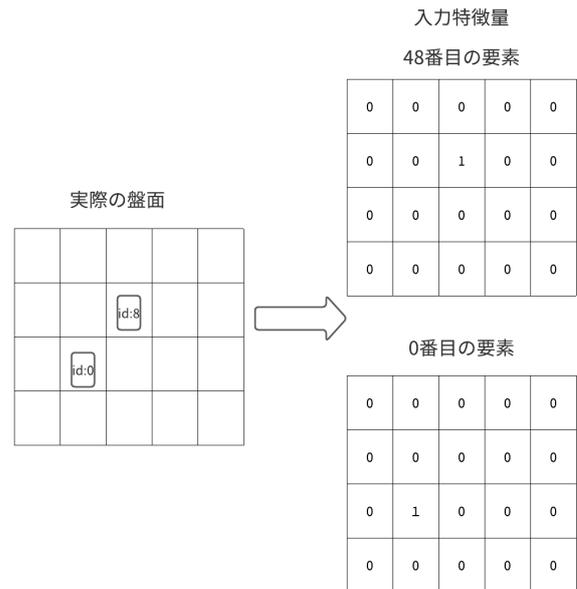


図 2 盤面と特徴量の対応

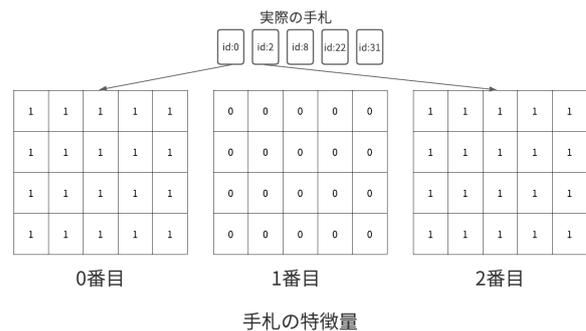


図 3 手札と特徴量の対応

カードの位置は 4 行 5 列のマスにカードが存在するならば 1、そうでなければ 0 とする。カードが手札に存在するならば 4 行 5 列の全てが 1、そうでなければ 0 とする。

このルールではカードはデッキ、手札、場、墓地のいずれかに存在し、カードを使用する際はそのカードを場に出し、カードの使用を宣言する。各プレイヤーは場、墓地、自身の手札に存在するカードを確認できる。本プログラムでは使用済みのカードを再利用するカードを採用しておらず、ルールの簡素さから相手の場のカードより高いステータスを持つカードを使用し戦闘によって相手の LP を 0 にするように行動するのが合理的であると考えられる。このような観点からカードゲーム対戦 AI への入力特徴量は自分と相手のカード 80 枚の位置を示す 4 行 5 列の特徴量を 80、自分の手札を示す 4 行 5 列の特徴量を 40 の合計 120 個の要素を持つ配列とする。

<sup>1</sup><https://ja.wikipedia.org/wiki/遊 戯 王 オフィシャルカードゲーム>

## 5.5 棋譜生成システムの実装

カードゲームの棋譜データのデータ拡張を行うために棋譜生成システムを実装する。SeqGAN の手法を参考にソースコードを変更しカードゲーム対戦 AI の教師データである「場と手札のベクトルの系列」と「何のカードをどこに出したかの ID の系列」のデータ生成を試みる。

ID の系列と棋譜データの生成をそれぞれ行った。SeqGAN では形態素解析した自然言語に ID を割り当て、データの生成を行っている。データ生成を行うにあたって生成するデータに対応したハイパラメータを設定しデータを生成した。しかし生成されたデータは両者とも本研究のカードゲーム特有の時系列に沿っているとはいえルールに則ったものでなく、十分な成果が得られなかった。

## 5.6 カードゲーム対戦 AI の実行

カードゲームプログラムを実行させカードゲーム対戦 AI の妥当性を評価した。カードゲームプログラムを自動対戦させ、カードを出すときの場の状況と出したカードの ID の系列を記録したものをデータセットとして用いる。自動対戦では現在の手札で最もステータスの高いカードを優先的に使用するよう設定した。場の状況を入力値、ID の系列を目標値としてデータセットを作成する。カードゲーム対戦 AI に場の状況を入力させ、出力された手と ID との誤差からパラメータを調整する。これによって入力された場に対して最適な一手を学習させていく。

## 6 考察

### 6.1 カードゲームの棋譜データのエンコード手法の考察

カードゲームの棋譜のベクトル化の方法の妥当性について考察する。100 戦学習したカードゲーム対戦 AI と 300 戦学習したカードゲーム対戦 AI をそれぞれ実行し、カードゲーム対戦 AI の出力する手を確認した。両者の実行結果には共通して以下の問題点が見られた。

- 手札に存在しないカードを出力する
- 同一の値の出力を続ける

問題の原因としてルールの簡素さが考えられる。簡略化したルールではステータスの最も高いカードを場に出し攻撃を行うのがゲームに勝利するのに合理的な手段になっている。このことから自動対戦では目標値に偏りが生じる可能性がある。

### 6.2 棋譜生成システムの考察

本研究では対戦カードゲーム AI の教師データの中で目標値となるデータは生成できたが入力値となるデータは生成できなかった。原因として SeqGAN と本研究の棋譜生成システムとでデータの形態や特徴が異なり、それを十分に吸収することのできる GAN アーキテクチャとはならなかったことが考えている。

仮に、棋譜生成システムの実装において妥当な棋譜データの生成が可能になった場合には、以下の観点から評価を行うことができる。

- 生成した棋譜のみを用いた場合のカードゲーム対戦 AI の勝率

### ● 棋譜生成システムの学習状況

棋譜生成システムが生成した棋譜の妥当性を評価するために生成した棋譜のみを用いてカードゲーム対戦 AI の学習を行う。棋譜生成システムの学習状況からシステムそのものの性能を評価する。

## 7 おわりに

対戦ゲームにおける AI は熟練者の振る舞いを再現したり、利用者に応じた難易度での対戦の実現などの用途での研究が進められている。対戦 AI は棋譜を用いて学習し、なかでもカードゲームは考えうる棋譜の種類が多岐にわたる。カードゲームの対戦 AI の教師データとなるカードゲームの棋譜を人力で収集するには多大な労力を要する。

カードゲームの棋譜を機械学習で学習できるようにカードゲームの棋譜のベクトル化方法を提案する。離散型時系列データのデータ拡張手法として棋譜生成システムを提案する。評価指標を定義し、ベクトル化手法とデータ拡張手法の妥当性を確認する。

本研究では棋譜生成システムを設計し、カードゲーム対戦 AI とカードゲームプログラムを実装した。棋譜生成システムの生成した棋譜でカードゲーム対戦 AI を学習させ、カードゲームプログラムで実行することで生成した棋譜の妥当性を評価する。結果として棋譜生成システム、カードゲーム対戦 AI いずれについても、棋譜の生成や対戦は行えるものの、十分な結果は得られなかった。

今後は棋譜生成システムやカードゲーム対戦 AI にとって最適になるようにルール、データを改善する。棋譜生成システムによって棋譜データを大量に生成し、カードゲーム対戦 AI の実現を支援を目指す。

## 参考文献

- [1] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. vanden Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, D. Hassabis: "Mastering the Game of Go with Deep Neural Networks and Tree Search", Nature 529, pp.484-489, 2016.
- [2] 竹内大輔, 野呂昌満, 沢田篤史: "ゲーム対戦戦略をプレイヤー習熟度へ適応させる機械学習機構の設計", 第 28 回ソフトウェア工学の基礎ワークショップ (FOSE2021), pp. 133-138, 2021.
- [3] Lantao Yu, Weinan Zhang, Jun Wang, Yong Yu: "SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient", The Thirty-First AAAI Conference on Artificial Intelligence, pp. 2852-2858, 2017.
- [4] AlphaGo で学ぶ強化学習 次の一手を予測する将棋 AI を作ってみよう - OPTiM TECH BLOG, <https://tech-blog.optim.co.jp/entry/2021/12/02/100000>.