Vol.2024-GI-51 No.33
2024/3/9

複数種類の戦略を持つプレイヤが混在する不完全情報ゲームにおける相手プレイヤの推定にむけて

木島花蓮 [1] 横山大作 [1]

概要：本研究では、多様な戦略を有するプレイヤが混在する多人数不完全情報ゲームにおいて、各プレイヤの戦略を推定可能なエージェントの開発を目的とした。研究対象として、各プレイヤが陣営ごとに戦略をもって行動する、シャドウレイダーズ [2] という不完全情報ゲームを用いた。深層強化学習を用いて、ゲーム環境と各プレイヤの行動から学習を行い、プレイヤの陣営や戦略を推定する手法を提案した。実験の結果、提案手法で陣営の差を推定することに成功したが、細かい戦略の違いについてはまだ課題が残ることも確認された。

Towards estimation of opponent players in incomplete information games with multiple types of strategies.

KAREN KIJIMA[1] DAISAKU YOKOYAMA[1]

1. はじめに

多人数不完全情報ゲームと呼ばれるゲームの中には、隠された役職（陣営）があり、それぞれが相手の役職を推定しながら、また場合によっては自分の正体を隠しながら、進めていくゲームがある。自分がゲームの全ての情報を得ることができない不完全情報ゲームは、社会全ての情報を得ることができない実社会と通じる物があると考えられ、中でも、相手が嘘をついていることや隠し事をしているという状況は実社会において実際に多く存在する。そのため、ゲーム上で行動等から隠された陣営や正体を見極めることは、ゲームのみならず、現実社会でも同様に隠された嘘や正体を見抜くことで問題の解決につながると考えられる。

このようなゲームの既存研究として、人狼ゲームの研究が多く挙げられる。しかし、人狼ゲームはプレイヤの発言をもとに推定を行うゲームであり、この発言を分類してそれをもとに陣営を考える必要もあり、コンピュータプレイヤが推定以外の複数の要素を必要とする。また推定ができたとしても、直接的に相手を攻撃することはできず、勝率に直結するとは限らない為、推定手法の評価が困難である。

そこで、本論文では、発言を推定に加味する必要がなく、また推定結果から直接的に攻撃もできる、シャドウレイダーズというゲームを研究対象とすることとし、より評価が出やすい環境で、相手の戦略を推定する手法を検証することを目指す。このゲームを対象としたゲーム環境を構築し、他プレイヤの行動や環境情報を用いた陣営推定モデルと、その推定結果をもとに行動を行うエージェントを作成し、性能評価を行った。

2. 研究対象

2.1 シャドウレイダーズとは

シャドウレイダーズは、4 人以上で行う正体隠匿型不完全情報ゲームである。各プレイヤは、ゲーム開始時に、それぞれの属性、勝利条件をもつキャラクターを与えられる。中でも属性は、シャドウ、レイダーという敵対する陣営と、第三陣営でありそれぞれ個別の勝利条件をもつシチズンからなる。

ゲームは移動、移動先のエリアボードの指示による行動、他プレイヤへの攻撃の 3 ステップをそれぞれのプレイヤが順番に行っていくことで進められる。シャドウ、レイダーはそれぞれ、勝利条件であるお互いの陣営のプレイヤの追放を目指し、シチズンはそれぞれの勝利条件の達成を目指す。

2.2 本研究でのシャドウレイダーズのルール

シャドウレイダーズはランダム性の高いゲームであり、例えば攻撃のダメージ量をダイスの目で決める等、推定が正しく行えてもダメージが 0 となってしまうこともある。そのため、より評価がしやすい環境になるよう、本研究ではゲームのルールを一部改変し、ランダム性を下げた状態でプレイを行う。

プレイ人数は 5 人とし、シャドウ陣営 2 人、レイダー陣営 2 人、シチズン陣営 1 人で行う。シチズン陣営には、ゲーム終了時に自身が生き残っていることが勝利条件である

1 明治大学大学院
2 http://www.cosaic.co.jp/games/sr.html

キャラクターと、自身が一番初めに脱落することが勝利条件であるキャラクターのどちらかが選ばれる。

各プレイヤはまずサイコロを二つ振り、6 つのエリアがあるエリアボード(表 1)の中で、出目に沿った番号のエリア(表 1 の番号に値する)に移動する。エリアごとにカードを引く、またはプレイヤへの攻撃や回復を行う等の指示があり、プレイヤはこれに従う。ここでカードとは、攻撃や回復を行うカードである「黒のカード」「白のカード」と、他のプレイヤの陣営を絞れる「推理カード」の 3 種類がある。推理カードを引いたプレイヤは他の任意のプレイヤにカードを渡し、そのプレイヤはそこに書かれているキャラクターの陣営またはキャラクターが自分のキャラクターと一致している場合、そのカードに書かれている指示に従う。一致していない場合、何も起こらないことを宣言する。

表 1　エリアボード

番号	名称	詳細
2，3	探偵事務所	推理カードを引く
4,5	ブラックミスト地区	推理カード、白のカード、黒のカードのいずれかを引く
6	大聖堂	白のカードを引く
7	地下通路	黒のカードを引く
8	市庁舎	プレイヤの内 1 人を選択し、2 ダメージ与える、又は 1 回復する
9	オリバーの隠れ家	他プレイヤの任意の装備カード(※)を 1 枚奪う

※白、又は黒のカードに存在する、所持している間書かれた効果が永続するカード

次に、自分以外のプレイヤの 1 人を必ず攻撃する。ここでの攻撃は 1 ダメージとし、最終的に HP ゼロになったものは脱落となる。脱落したプレイヤは自身のキャラクターカードを全てのプレイヤに開示する。

いずれかのプレイヤが自身の勝利条件を満たした場合、その時点でゲームは終了とし、ゲーム終了時点で勝利条件を満たしているプレイヤは、脱落したプレイヤを含めて全員勝利となる。

3. 関連研究

隠された役職のあるゲームにおいて役職や陣営の推定手法を検討した先行研究は、その大半が人狼ゲームを対象とした研究であり、例として、大川らの深層学習を用いた人狼の役職の推定[1]や、福田らの人狼ゲームにおける深層強化学習を用いたエージェント[2]などがある。これらの研究では、発言内容や占いの数、プレイヤの生死等を入力として深層学習を行っており、これらの学習に使用するデータは人狼知能大会でのゲームのログから作成されている。

しかし、本研究は人狼等の研究のさかんなゲームではなく、プレイデータの少ないゲームを対象としており、データの不足が懸念される。そこで、本研究ではデータログを入力とした教師あり学習ではなく、深層強化学習を用いた推定手法を検討した。

4. 提案手法

本研究では、5 人プレイのシャドウレイダーズにおいて、陣営推定モデルとそのモデルを搭載したエージェントを提案する。

4.1 ゲームの基本実装

本研究では、他プレイヤの行動から陣営を推定しながらシャドウレイダーズをプレイすることができるのかを検証することが目的である。そのため、大まかなプレイはルールベースで作り、陣営の推定のみを学習可能なフレームワークとし、深層強化学習でこの部分の最適化を行うこととした。これにより陣営を推定したうえでのプレイができるかどうか検証することが今回の目標となる。

シャドウレイダーズのプレイ部分の実装は OpenAIGym を用いて実装し、主にゲーム自体の進行を進める MyEnv クラスと、各プレイヤの情報を保持し、MyEnv クラスから渡される状態をもとにどのように行動するかを決める Player クラスに分かれ、図 1 のような形で実装する。MyEnv クラス内では 5 人のプレイヤが順番に移動、行動、攻撃を行っていく。ルール上、移動はランダムに行われ、移動したエリアボードでは Player クラスから渡される任意の行動を行い、同様に任意の攻撃対象を決めて攻撃を行う。また、Player クラスで保持するプレイヤの情報は、プレイヤ自身のキャラクター情報、プレイヤ自身が現在いるエリア、ダメージ量、推理カードから得た他プレイヤの陣営情報である。Player クラスでのプレイヤの行動決定は後節で述べる。

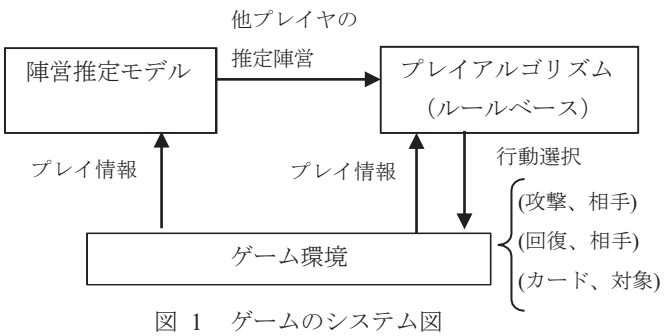

図 1　ゲームのシステム図

4.2 行動決定方法

各プレイヤの陣営ごとの行動選択基準として、人間がプレイする際に多くの場合に取るであろう行動に設定した。

主にシャドウ、レイダー陣営はお互いを敵陣営として、お互いの陣営の殲滅を目的とした行動を取る。また、シチズン陣営は、勝利条件が生き残ることである場合、どちらか弱った陣営の脱落を目指して行動し、勝利条件が最初に脱落することである場合、出来るだけどのプレイヤも脱落しないように行動する。

Player クラスにおいてプレイヤの行動決定を行うとき、プレイヤの意思決定が関わる場合と、ランダムに行動が選ばれる場合がある。プレイヤの意思決定が絡んでくるのは、攻撃または回復の対象を選ぶ際と、ブラックミスト地区にて 3 種類のカードのいずれかを選ぶ際、また推理カードを渡す対象を選ぶ際の 3 種類がある。

攻撃対象を選ぶ際は、それぞれのコンピュータプレイヤの推定結果から一番敵陣営らしいプレイヤを選択する。回復対象を選ぶ際には、自身も回復対象の候補に含まれる場合のみ、自身のＨＰが半分以下の際は自身を回復するとし、もし自身のＨＰが半分以上の場合は、推定結果から一番味方らしいプレイヤと自身のどちらかで残りのＨＰが少ない方を対象とする。自身が対象に含まれない場合は、一番味方らしいプレイヤを回復対象とする。ここで一番味方らしいプレイヤのみを見るのは、5 人プレイの際は、自身の味方陣営は 1 人のみのためである。

また、3 種類のカードからいずれかのカードを選ぶ際は、まだ誰の陣営も推定できていない場合、推理カードを選ぶ。それ以外の場合では、シャドウ陣営の場合は 70%の確率で黒のカード、30%の確率で白のカードを選び、レイダーの場合は 70%の確率で 30%の確率で白のカードを選ぶ。これは、白のカードにはレイダーに有利でシャドウに不利なカードが、黒のカードにはその逆のカードがいくつか入っており、実際のプレイの際には、それぞれレイダーは白、シャドウは黒のカードを選ぶことの方が多い傾向にあるためである。ただし、それぞれのカードがすでに残っていなかった場合は、推理カードを優先して引き、これもなかった場合、残りの別のカードを引く。

推理カードを渡す対象を選ぶ際には、まだ推定できていないプレイヤがいた場合そのプレイヤに渡し、すべてのプレイヤの推定に確証を持っている場合は、ダメージを与える可能性のあるカードなら敵に、回復する可能性のあるカードなら味方に渡すようにする。

4.3 陣営推定

学習を行うプレイヤはシチズン陣営以外とし、正しく敵味方とそれ以外を判別できるかどうかを判断する。また、シチズン陣営は敵味方以外のノイズのプレイヤとして用意し、ランダムな行動を取るプレイヤとした。

4.3.1 陣営推定モデル

陣営推定モデルでは、Deep Deterministic Policy Gradient[3]を使用し、各プレイヤについて、表 2 の 3 つの特徴を 1 つのベクトルとして結合して入力とし、そのプレイヤのそれぞれの陣営である確率をソフトマックス関数にかけて出力する。また、報酬は各エピソードの終了時点での推定精度とし、出力のうち実際のプレイヤの陣営らしさを計算しており、実際の陣営の確率をどれだけ高く出せているかを報酬として与えている。つまり、出力の値のうち実際の陣営の確率を出している値をプレイヤ 4 人分取り出し、その数の合計を 4 で割った数値が報酬となる。

表 2　プレイヤ X の特徴

特徴	詳細
攻撃対象	プレイヤ X が攻撃を行った際の対象のプレイヤ
回復対象	(プレイヤX が回復を行う場合のみ)プレイヤX が回復を行った際の対象のプレイヤ
推理カードの情報	エージェントが推理カードから得られたプレイヤ X の情報
3 種類のうち選択したカード	(プレイヤX がカードの選択を行うエリアに移動した場合のみ)プレイヤX が選択したカード

Deep Deterministic Policy Gradient の探索の際の基本的なハイパーパラメータはkeras-rl の標準的な実装の通りとし、探索の際に与えるノイズには平均回帰過程を用い、最適化アルゴリズムには Adam[4]を使用した。モデルの入力次元は 20、出力次元は 12 である。Actor Network、Critic Network 共に隠れ層は 3 つの 128 ユニットの Dense 層であり、活性化関数には ReLU を用いている。また、Actor Network の出力層には 3 つごとにグループ化してソフトマックス関数にかけた Lambda レイヤを用いており、Critic Network の出力層の活性化関数には Linear を用いている。

3 万エピソード学習を行った結果、以下の図のように学習が収束した。ここで、横軸は 50 ごとのエピソード数、縦軸は報酬である。

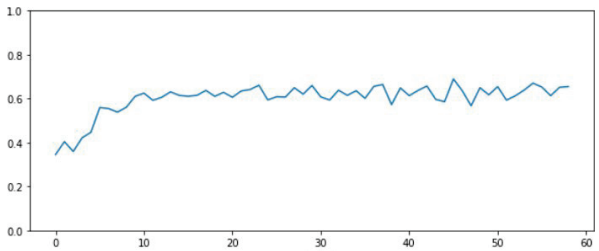

図 2　陣営推定モデルの episode-reward

4.3.2 結果

提案手法を用いたコンピュータプレイヤ 1 人と、ヒューリスティックなコンピュータプレイヤ 4 人でシャドウレイダーズをプレイする。

ヒューリスティックなコンピュータプレイヤは推理カードの情報のみから推定を行い、推定ができている場合はその内容から攻撃や回復対象を選択し、出来ない場合はランダムに選択する。

1 ゲームの終わりにどれだけ正しく陣営を推定できているかを、陣営推定モデルの出力結果から、実際の陣営の確率が 75%を超えた場合推定できていると仮定し、推定精度の評価を行った。比較対象として、ヒューリスティックなコンピュータプレイヤの推定精度、ランダムなコンピュータプレイヤ（適当に推定の確率を出すプレイヤ）での推定精度も測定する。それぞれ 1 万回試行した結果、各推定精度は以下の表 3 の通りとなった。

表 3　推定精度

	提案手法	ヒューリスティック	ランダム
推定精度(%)	69.82	54.43	37.09

また、陣営ごとの推定精度としては、以下の表 4 の鳥となった。

表 4　役職ごとの推定精度

	敵陣営	味方陣営	シチズン
推定精度(%)	66.98	72.24	62.85

どの陣営も大きな差はなく推定できているが、味方陣営が特に推定精度が高くなった。これは、味方陣営は自分ともう一人であり、1 人のみの推定であるためであると考えられ、同様に 1 人のみの推定であるシチズンの推定精度が低いのは、シチズンにランダム性がありどちらかの陣営と推定してしまう場合があるためであると考えられる。

提案手法を用いた場合の推定精度はヒューリスティック、ランダムなプレイヤより高くなっており、提案した陣営推定モデルが有効であるといえる。

また、提案手法を用いたコンピュータプレイヤの勝率と、ヒューリスティック、ランダムなコンピュータプレイヤの勝率、また推定が完璧である場合の勝率は、同様に 1 万回試行した結果、以下の表 5 の通りとなった。

表 5　勝率

	提案手法	ヒューリスティック	ランダム	推定 が 100%
勝率(%)	59.25	53.37	44.13	68.28

推定精度が高いものが勝率は順当に高くなっており、シャドウレイダーズにおいて推定精度の高さは勝率に直結するといえる。これは、推定が上手くいっている場合は攻撃で敵だと推定している一人のプレイヤを攻撃し続け、また自身や味方陣営が死なないように回復を行いながらプレイできるようになっていたためであるといえる。

4.4 戦略推定

プレイヤが陣営のみではなく、その勝利条件まで推定できることを目標として学習を行った。陣営の推定とは異なり、各プレイヤの行動決定のパターンを見極めることが目的ではあるが、シャドウ、レイダー陣営はどのプレイヤも敵陣営の殲滅が目的であるため、結果としてシャドウ、レイダー等の陣営の推定も同時に行えることとなる。また、陣営の推定と大きく異なる点として、シチズンは個々の勝利条件を持つため、行動決定方法が違うことがあげられ、この部分の推定が主な目標となる。

4.4.1 戦略推定モデル

前節の陣営推定モデルでの出力の確率を、敵陣営、味方陣営、シチズン 1、シチズン 2 として出力する。ここで市民 1 は自身が生き残ることが勝利条件のキャラクター、市民 2 は自身が最初に脱落することが勝利条件のキャラクターとする。

他の条件は全て同様で、3 万エピソード学習を行った結果、以下の図 3 の通り、学習は成功しなかった。

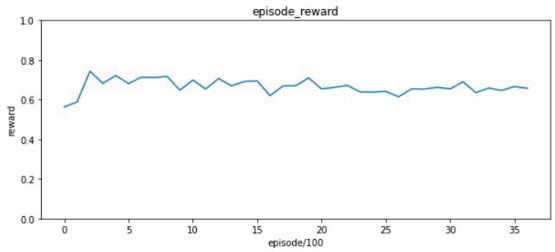

図 3　戦略推定モデルの episode-reward

結果として、生き残ることが勝利条件であるシチズンプレイヤを敵または味方陣営と誤認していることが多いことがわかった。これは、どちらか一方の陣営の脱落を手助けする傾向であるためだと考えられる。

5. まとめ

深層強化学習を用いた陣営の推定自体は70%程度の精度で可能であったものの、まだ改善の余地があると考えられる。また、それぞれの勝利条件ごとの戦略の違いについては学習が上手くいかなかった。
今後の課題として、陣営の推定の精度を上げることや、戦略の推定を行えるようにモデルの改良を検討するほか、ヒューリスティックなプレイヤ同士でのゲームのログを取得し、そこからの学習を行うことも検討したい。

参考文献

[1] 大川喜聖、吉仲亮、篠原歩: 深層学習を用いて役職推定を行う人狼知能エージェントの開発,第 22 回ゲームプログラミングワークショップ, pp.50-55, 2017
[2] 福田 宗理, 穴田 一:15 人人狼ゲームにおける会話情報による役職推定,情報処理学会第 82 回全国大会, pp95-96, 2020
[3] Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, Daan Wierstra: Continuous control with deep reinforcement learning. ICLR (Poster) 2016
[4] Diederik P. Kingma, Jimmy Ba: Adam: A Method for Stochastic Optimization. ICLR (Poster) 2015