

情報工学実験 2 10/18 課題

タイトル

不完全情報ゲームにおける強化学習を用いた戦略の構築とその分析 [1]

著者

阿部慎太郎, 竹川高志

何に関する研究か

プレイヤーに与えられる情報が部分的である不完全情報ゲームでは, 駆け引きが生じるため必勝手は存在しない. また, 不完全情報ゲームの中でもプロ選手が生まれていることから無数にある戦略の中でも総合的な優劣があると考えられる.

この研究では, このような不完全情報ゲームにおいて平均的に勝つことができる戦略の構築を目指している. 実験の際に使用する不完全情報ゲームとしては, 5 枚のカード ver の「ハゲタカのえじき」を採用している. 図 1 に「ハゲタカのえじき」のゲームの流れを示す.

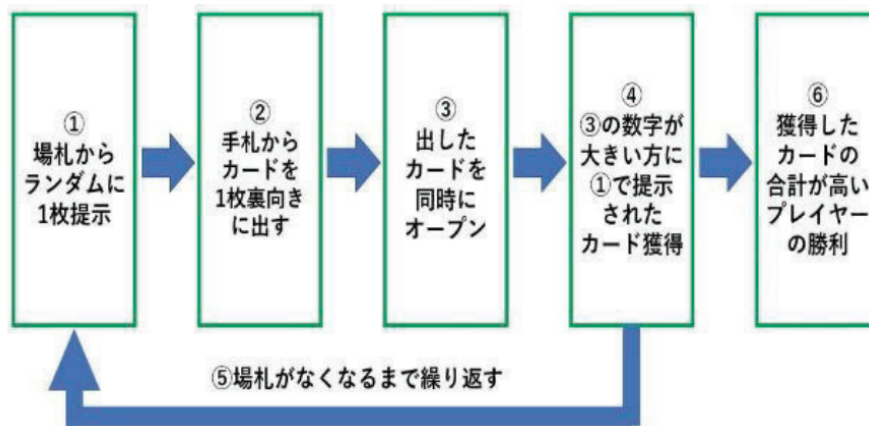


図 1: 「ハゲタカのえじき」のゲームの流れ

実験内容

この章がないと筆者の主張や興味深い点のが伝わりにくいと考えたため追加した.

強化学習を用いて特定の対戦相手との対戦から counter 戦略を作成する. 不完全情報ゲームには平均的に勝つことのできる強い戦略が存在するという仮定に基づき, counter 戦略の対戦を繰り返すことでより強い戦略の構築を目指す.

実験の際には, 手札からランダムにカードを出す random 戦略, バフだと同じカードを数字から出す固定戦略, DQN を用いて相手との対戦から学習を行いカウンターとなる行動を選択する counter 戦略, 複数の戦略の

中から対戦ごとに強い戦略を多く選択するように戦略を切り替える切り替え戦略の4つの戦略を設定した。また、本研究では counter 戦略 vs random 戦略の対戦, counter 戦略 vs 固定戦略の対戦を行っている。

counter 戦略 vs random 戦略

counter 戦略 vs random 戦略の対戦では、ランダム戦略との対戦を繰り返して学習した counter_random 戦略を構築した。

counter 戦略 vs 固定戦略

counter 戦略 vs 固定戦略の対戦では、まず、counter 戦略 vs 固定戦略の対戦から学習し、counter_1 戦略を作成した。同様に counter_1 戦略の対戦から学習した counter_2 戦略を作成し、固定、counter_1、counter_2 の3つの戦略でグループを構築し、グループの総当たりから勝率の割合を計算し、切り替え戦略_1を作成する。そして、counter 戦略と切り替え戦略_1の対戦から学習した counter_mix1 戦略をグループに追加する。この流れを図2のように繰り返して counter_mix7 戦略まで作成した。



図 2: 固定戦略を基にしたグループの作成方法

著者が主張している点

counter 戦略 vs random 戦略

counter_random 戦略は random 戦略に対して約 88% 場札と同じ数字のカードを手札から出す傾向があることが判明した。すなわち random 戦略には固定戦略が有効であることが判明した。

counter 戦略 vs 固定戦略

counter 戦略に関しては 1 つ前の戦略にはどの戦略も 90% 以上の勝率になった。また、最も新しい世代の戦略である counter_mix7 戦略はグループ内のすべての戦略に対して勝率が 60% 以上となった。

また、この counter_mix7 戦略の行動における各ゲーム情報の重要度を決定木分析で行った。各ゲーム情報は自分の手札、相手の手札、残りの場札、得点差であり、結果として counter_mix7 戦略は得点差と残りの場札に関する情報の重要度が高く、各プレイヤーの手札に関する情報の重要度は低くなった。

結論と今後の課題

このように、対戦相手に合わせた強化学習により、対戦相手に平均的に勝つことのできるゲーム AI の構築が可能だと言える。

本研究では予め対戦相手の戦略を設定していたが、ゲームの途中で戦略を変更する、あえて弱い行動をして駆け引きに持ち込むなど複雑な戦略相手に勝利するためには情報の集約や対戦相手の特徴の把握などの学習の際の工夫が必要である。人間との対戦から効率よく学習を行い、平均的に勝つことのできる戦略を探索していく。

興味深い点

counter_mix7 戦略の戦略決定の情報

不完全情報ゲーム、特に本実験の「ハゲタカのえじき」では、相手の手札というプレイヤーに与えられない情報が存在する。決定木分析により得られた counter_mix7 戦略の戦略決定において、他プレイヤーの手札という本来得られない情報の重要度は低く、得点差と残りの場札といった盤面から取得可能な情報の重要度は高くなっていた。この情報の重み付けが学習の際に最適化された結果であるかどうかは本論文では明かされていないが、counter_mix7 戦略が盤面の情報から戦略を組み立てて一定の成果を上げていることが面白かった。

切り替え戦略を用いた counter_mix 戦略の作成

counter_2 戦略を作った後に図 2 のように戦略を同一グループに入れ、切り替え戦略を作成し学習を行っていた。単純に counter_2 戦略を作成後に counter_2 戦略と固定戦略を学習させて counter_3 戦略を作成する方法では恐らく過学習の問題が起るため本研究はこのような方法にしたと考えたが、興味深い実験方法だなと感じた。

この次に読むべき資料

自分が理解できてなかったキーワード

- AlphaGo
<https://www.nature.com/articles/nature16961>
- DeepBlue
<https://cir.nii.ac.jp/crid/1050845762805286016>
- 決定木分析
<https://qiita.com/3000manJPY/items/ef7495960f472ec14377>

より複雑なゲームの学習

- AlphaStar(starcraft2 を解く AI)
<https://www.deepmind.com/blog/alphastar-mastering-the-real-time-strategy-game-starcraft-ii>

不完全情報ゲームの深層強化学習応用例

- 逆転オセロニアへの応用
<https://www.ipsj.or.jp/dp/contents/publication/38/S1002-S05.html>

参考文献

- [1] 阿部慎太郎, 竹川高志. 不完全情報ゲームにおける強化学習を用いた戦略の構築とその分析. 人工知能学会全国大会論文集, Vol. JSAI2022, pp. 4E1GS201–4E1GS201, 2022.