

ボードゲームを対象としたブラフへの AI 適用検討

河地 智哉[†]尾崎 敦夫[‡]大阪工業大学 情報科学部 コンピュータ科学科[†] (現, 情報知能学科[‡])

1. はじめに

近年、AI をボードゲームへ適用した事例が多数報告されている[1]。多種あるボードゲームの中で、ポーカー[2]は、心理戦を特徴とし、自分の手役を強くみせるためにブラフを使う。ポーカーでのブラフとは、自分の役が弱いにも関わらずビット(チップを出すこと)を多くして自分の役を強く見せることである。このブラフにより相手を欺き、勝負から離脱させること(ドロップ)もできる。

本稿では、ポーカーを対象とした心理的な駆け引き(ブラフ)において、勝率向上を図るための AI 適用法を提案する。

2. 対象ゲーム「ポーカー」の概要

ここでは、クローズドポーカーを対象に説明する。使用するのはトランプ 52 枚とチップである。最初に各プレイヤーは参加料として 1 枚のチップを場に出す。そして、それぞれ 5 枚ずつのカードが伏せて配られる。残りは中央の場に裏向きに置く。各プレイヤーは配られた手札を見て、順番に「ビット(チップを出すこと)」または「パス」を宣言する。誰か 1 人がビットしたら、それ以後の人はパスできない。最初にビッドする人は、チップの枚数を宣言して場に出す。そして、次の人からは、先のプレイヤーと同数のチップを出す(コール)か、多いチップを出す(レイズ)かしてゲームを続行するか、ドロップするかを選択する。レイズする人は、チップ枚数を宣言する必要がある。最高のビッドに対して、誰もレイズをしなければビットは終了する。ゲームに残った人が 3 人以上いた場合は、順番に手札を中央の場のカードと交換する。その後、手札を公開し、他のプレイヤーと役を比較して、その中で最強の役を持ったプレイヤーが勝ちとなり、場にあるチップを全部獲得する。本来はこれを何巡か繰り返すが、初期の検討では、まずは手札の交換を 1 回とし、プレイヤーが 3 人の場合において、学習する状況を考える。

3. 提案手法

従来の Deep Q-Network[3]では報酬を 1 つ与えるが、ここでは異種の報酬を設定することによって勝敗だけでなく、相手を騙すことに対しても評価することを考える。具体的には、報酬は相手を騙せた時と相手に勝った時の 2 つの状況に対して与えることとした。ここで“相手を騙すこと”とは、最後に手札を公開した際に他のプレイヤーが自分よりも良い手にも関わらずドロップしたことを指すこととする。なお、報酬はそれぞれ r_a と r_b とし、相手を騙すことができた時と勝負に勝つことができた時で違う値に設定することとした(図 1 参照)。そして、

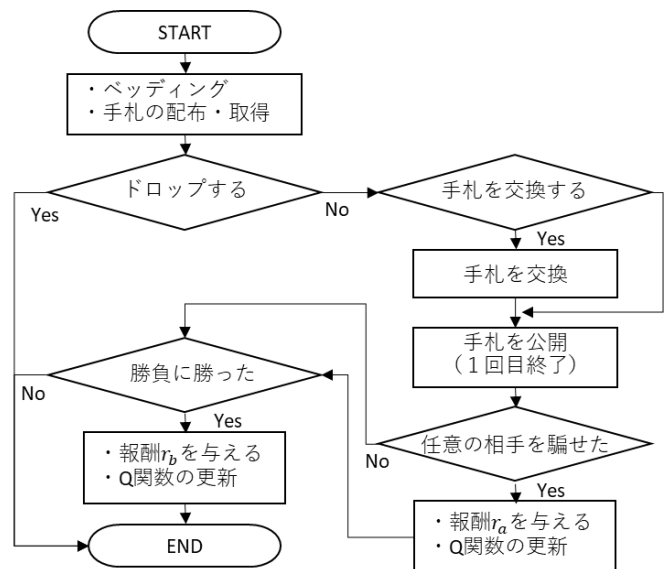


図 1. 提案手法による処理概要

手札を公開した後、勝敗を確認し報酬に応じて価値関数 Q を更新する。また、AI は価値関数 Q をもとに手札を交換するかどうかなどの行動を判断することとする。

4. 期待する効果

本提案手法は、ゲームの勝敗だけではなく相手を騙すことができたかどうかに対しても報酬を設定するため、今までと違ったブラフの実現が期待できる。これは、ポーカーなどのゲームだけではなく人間同士の交渉などにも使える可能性を秘めており、例えば、人間の代わりに交渉を行うような機械の実現も期待できると考える。

5. 今後の課題

本稿では、AI 学習での報酬の設定方法に関する手法を提案したが、これに加えて、モンテカルロ木探索などで先読みする処理も必須となる。また、弱い役で勝ち続けてしまうと偏った学習になる恐れがあり、役に対する報酬の細分化なども今後の課題となる。その他、手札の交換を複数回行う場合を考慮した AI 適用法など、様々な状況を考慮した検討を今後行う予定である。

参考文献

- [1] 伊庭斉志, “ゲーム AI と深層学習,” オーム社, 2019.8.25.
- [2] Nintendo, “遊び方:ポーカー,” https://www.nintendo.co.jp/others/playing_cards/howtoplay/poker/index.html
- [3] 小川雄太郎, “つくりながら学ぶ! 深層強化学習 PyTorch による実践プログラミング,” (社)マイナビ出版, 2018