

相手の抽象化による多人数ポーカーでの戦略の決定

古居 敬大^{1,a)} 浦 晃^{1,b)} 三輪 誠^{2,c)} 鶴岡 慶雅^{1,d)} 近山 隆^{1,e)}

概要：3人以上で行われるポーカーのような多人数ゲームにおいて、相手プレイヤーの人数が多少変わったとしても有効な戦略の性質は変わらないことが予想される．そこで本稿では相手プレイヤーの行動を削減・抽象化を行うことで、多くのプレイヤーが存在するゲームで実際のゲームより少ない人数のゲームを想定し、少人数ゲームで学習した戦略を適用する手法を提案する．簡易的な多人数のポーカーでの実験の結果、特に多人数でのゲームで十分に学習ができていない場合は提案手法が有効になり大きな報酬が得られることがわかった．

Computing Strategies for Multiplayer Poker Games by Abstraction of Opponents

KEITA FURUI^{1,a)} AKIRA URA^{1,b)} MAKOTO MIWA^{2,c)} YOSHIMASA TSURUOKA^{1,d)}
TAKASHI CHIKAYAMA^{1,e)}

Abstract: In multiplayer games like poker, it is assumed that small changes in the number of the players do not affect characteristics of the strategies so much. This paper proposes a method to apply strategies learned from games with a smaller number of players to games with a large number of players, by abstracting opponents' plays. Experiment results on simplified poker games show that the proposed method is effective especially when conventional players cannot learn playing strategies enough.

1. はじめに

トランプゲームのひとつであるポーカーは、手札が山札よりランダムに配られ、相手の手札が最後に見せ合うまでわからない、不確定不完全情報ゲームである．不確定不完全情報ゲームでは、考えられる現在状態の集合 (情報集合、図 1) について、確率的に良い報酬が得られ、かつ相手に自らの情報をあまり与えないような、ナッシュ均衡 [1] に基づく戦略 (ナッシュ均衡戦略) が取れることが望ましいとされている．

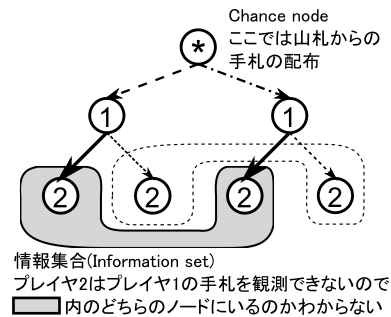


図 1 不完全情報ゲームと情報集合

¹ 東京大学大学院工学系研究科
Graduate School of Engineering, The University of Tokyo
² マンチェスター大学コンピュータ科学科
School of Computer Science, University of Manchester
a) furui@logos.t.u-tokyo.ac.jp
b) ura@logos.t.u-tokyo.ac.jp
c) makoto.miwa@manchester.ac.uk
d) tsuruoka@logos.t.u-tokyo.ac.jp
e) chikayama@logos.t.u-tokyo.ac.jp

ポーカーのような状態数の大きい展開型ゲームで、厳密なナッシュ均衡解を得ることは、空間計算量の観点から困難である．そのため、ナッシュ均衡戦略を行うためには、ゲーム状態の抽象化を行うことで近似的なナッシュ均衡戦略 (ϵ -ナッシュ均衡戦略) を求め、それを利用することとなる． ϵ -ナッシュ均衡戦略については、従来は線形計

画問題を解くことで得ていた [2]。しかし近年オンライン学習で注目されている Regret 最小化の概念を導入した、CounterFactual Regret minimization (CFR) [3] が提案された。これは従来の手法に比べ空間計算量的に優れており、より正確なナッシュ均衡の近似戦略を得られるようになった。その成果もあり、チャンピオンクラスの間人プレイヤとの 1 対 1 の対戦で勝ち越すまでになった [4]。

またポーカーは 2 人から 10 人程度までが参加可能な多人数ゲームでもある。多人数のポーカーはプレイヤの参加人数に対して、状態数が指数的に増加するため、従来の主な研究対象であった 2 人ポーカー以上に、ナッシュ均衡戦略の計算が難しくなる。そのため多人数ゲームでの戦略の決定は、かつては人数に依存しないがプログラマのゲームの知識や経験に頼ったルールベースに基づくアプローチ [5] で、近年では CFR 等の従来 2 人ゲームで行われてきた手法の 3 人ゲームへの適用といった人数を限定したアプローチ [6] での研究が主流であり、人数に依存せず ϵ -ナッシュ均衡戦略を適用する手法ほとんどなかった。

本稿では従来の CFR が扱ってきた人数 (2, 3 人) より多い人数が参加するゲームを対象とした、 ϵ -ナッシュ均衡戦略の適用手法を提案する。相手プレイヤの行動を削減・抽象化することで、ゲーム状態の構造や特徴を保持しつつもより少ない人数のゲームを想定することが可能となり、少人数のゲームでの ϵ -ナッシュ均衡戦略を適用することが可能となると考えている。

戦略の学習時間の問題や比較するプレイヤを生成するため、今回の実験ではポーカーのトイゲームであるルダックホールデム [7] を 6 人まで参加できるように拡張したゲームを用いる。

2. ポーカー

ポーカーはトランプを用いて行うゲームの一種である。基本的には最終的な手の役^{*1}の強さで勝敗を決めるゲームであるが、相手全員をフォールドさせゲームから降ろすことができれば、手札の強さによらず勝てるという特徴がある。他の相手プレイヤからは手札が見られないこともあり、手札が弱い時でも強気の行動を取るブラフなども有効に作用するゲームであるといえる。

2.1 テキサスホールデム

テキサスホールデム (Texas Hold'em) はポーカーの一種であり、世界的に人気のあるゲームである。

テキサスホールデムの特徴は複数あるベットラウンドと全プレイヤが共通して使用するコミュニティカードにある。

ベットラウンドは 1 回のゲームに 4 回あり、各ベットラウンドでは場のカードや相手の行動をもとにゲームを続行

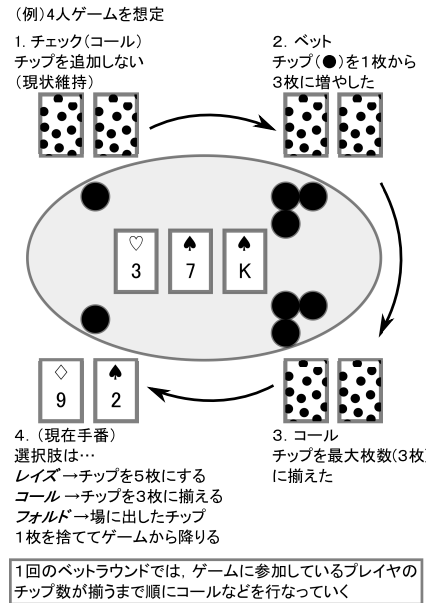


図 2 ベットラウンドでのプレイヤの行動

するかを決断を行う。ベットラウンド中の具体的な選択肢は以下の 3 つである。

- レイズ (ベット) : 賭けるチップを増やす。ルールによってはレイズの回数やチップの枚数に制限がある。
- コール (チェック) : 賭けに同意し相手にチップを揃える。
- フォールド : ゲームから降りる。賭けたチップは戻ってこない。

これらの行動の具体例は図 2 に示した。

コミュニティカードは各プレイヤ間で共有する、役を作るためのカードであり、ベットラウンド毎に枚数は増えていく。最後までゲームに参加していた各プレイヤはコミュニティカード 5 枚と手札 2 枚から 5 枚を選んで役を作り勝負を行うことになる。

このようにベットラウンドを重ねて賭けるチップを増やしていき、図 3 のように 4 回のベットラウンドを経て、最終的に手札を見せ合い勝敗を決定する。賭けられたチップは全て勝ったプレイヤのものとなる。

テキサスホールデムはプレイヤの行動順によって各プレイヤの得られる情報に差が生じる不完全情報ゲームである。概して前のプレイヤの情報が利用できるため、プレイ順序が遅い方が有利という特徴がある。そのため順序の早いプレイヤは出来るだけ自身の情報を明かさないように行動し、遅いプレイヤは早いプレイヤの行動から出来るだけ有効な情報を抽出し行動することが望ましい。

2.2 ルダックホールデム

ルダックホールデム (Leduc Hold'em) [7] は、カードの共有やベットラウンドといったテキサスホールデムの基本的な特徴を保持させながら、カードの種類や枚数を大きく

*1 ワンペア、ツーペアなど

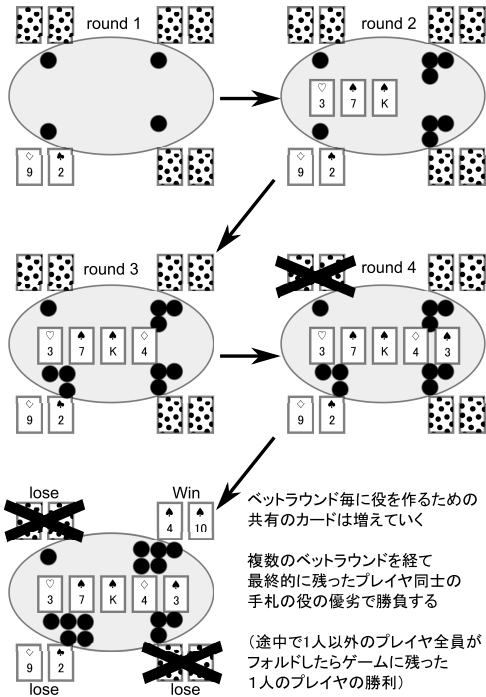


図 3 テキサスホールデムのゲームの一連の流れ

表 1 テキサスホールデムとルダックホールデムの比較

比較項目	Texas	Leduc
山札の枚数	52 枚	$2(n+1)$ 枚
手札	2 枚	1 枚
共有のカード	3~5 枚	1 枚
ベットラウンド	4 回	2 回
1 ラウンドのレイズ	4 回まで、等	2 回まで
役	ストレートやスリーカードなど	ワンペアとハイカードのみ
状態数	10^{18}	$\sim 10^n$

限定したトイゲームである。おもにテキサスホールデムの分析やコンピュータプレイヤの評価の目的で用いられている。

テキサスホールデムとの違いは表 1 のとおりである。ルダックホールデムでは 1 ゲームの参加費として 1 枚のチップを賭け、報酬として -13 から $+13(n-1)^2$ 枚のチップを得られるゲームである。

カードの組み合わせを除いた状態数は、プレイヤの人数が n 人であればおよそ 10^n 通りであり、最も簡単な 2 人ゲームでも状態数が 10^{18} あるテキサスホールデムより大きく簡略化されていることがわかる。

3. 関連研究

3.1 ナッシュ均衡

ナッシュ均衡 [1] はゲーム理論の非協力ゲームの基本的な解である。プレイヤ全員が独立に思考し利己的かつ合理的であるという仮定をおいたとき、どのプレイヤも 1 人では行動を変更することで現在より大きな利得を得られない、

*2 n はゲームへのプレイヤの参加人数

という均衡した状態を指す。基本的に非協力零和ゲームであるポーカーなどのゲームでは、ナッシュ均衡に基づく戦略は相手の戦略がわからない限り最善の戦略といえる。

一般的に均衡解の計算には、あらゆる状態での行動の組み合わせ（純戦略）のあらゆる場合をベクトルとし、それら純戦略の組み合わせの報酬を要素とするような行列（利得行列）を用いた計画問題を解き、純戦略の確率的な足しあわせである混合戦略を求めることになる [2]。

相手に自身の戦略を利用されないために、自身の情報を最大限与えないという特徴から、ポーカーのようなゲームでは弱いのに強気に見せるブラフといった行動もナッシュ均衡から確認できる*3。

3.2 ϵ -ナッシュ均衡戦略と

CounterFactual Regret minimization (CFR)

ポーカーのような展開型ゲームでは状態数が多くなるため、厳密なナッシュ均衡戦略を計算することが困難である。そのため計算が可能になるまでゲームの状態を抽象化・簡易化し、近似的なナッシュ均衡戦略 (ϵ -ナッシュ均衡戦略) の確率表を事前に求めてそれを利用することになる。

2007 年に Zinkevich らが提案した CFR は、従来の線形計画法を用いたナッシュ均衡の近似解法に比べ、空間計算量に優れた手法である。CFR によって扱える状態数が従来の 10^8 から 10^{12} 程度まで大きくなり、2 人レイズリミット・テキサスホールデムではプレイヤの行動の構造を抽象化せずに直接扱うことが可能になり、結果としてチャンピオンクラスの人間プレイヤに勝利できるまでになった強力な手法である [4]。

CFR はゲーム木を情報集合毎に集約した木 (information set tree) として表現し、各イテレーション毎に Regret と、それを利用した戦略を更新していく繰り返しアルゴリズムである。各情報集合ノード I では Regret の指標として以下のような R を定義する。

$$R_i^T(I, a) = \frac{1}{T} \sum_{t=1}^T \pi_{-i}^t(I) (u_i(\sigma^t|_{I \rightarrow a}, I) - u_i(\sigma^t, I)) \quad (1)$$

図 4 のように、ここで i は I での手番プレイヤ、 $\pi_{-i}^t(I)$ は戦略 σ^t を i 以外のプレイヤが利用した時のノード I へ到達する確率*4、 $u_{i,I}$ はノード I でのプレイヤ i の平均報酬であり、 $u_i(\sigma^t|_{I \rightarrow a}, I)$ はプレイヤ i がノード I で行動 a を行った場合の平均報酬である。 $u_i(\sigma^t|_{I \rightarrow a}, I) - u_i(\sigma^t, I)$ が負になる行動 a は平均より損をする行動で、正になる行動は平均より得をしてしまう (過去の行動で機会損失が生じている) 行動ということになる。そこで Regret の中から機会損失が生じている R が正の行動のみを式 2 によって抽出してそれを R^+ とし、式 3 のように次回イテレーションで

*3 One-card poker

<http://www.cs.cmu.edu/~ggordon/poker/>

*4 手番が i のノードでの遷移確率は 1 とする

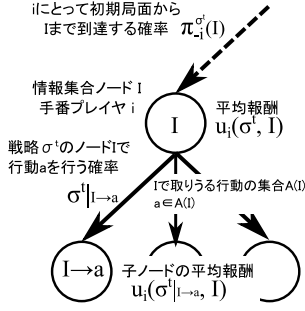


図 4 情報集合ノードとそこでの戦略と平均報酬

利用する戦略 $\sigma_i^{T+1}(I)(a)$ へ用いる.

$$R_i^{T,+}(I, a) = \max(R_i^T(I, a), 0) \quad (2)$$

$$\sigma_i^{T+1}(I)(a) = \begin{cases} \frac{R_i^{T,+}(I, a)}{\sum_{a \in A(I)} R_i^{T,+}(I, a)} & \text{if } \sum_{a \in A(I)} R_i^{T,+}(I, a) > 0 \\ \frac{1}{|A(I)|} & \text{otherwise.} \end{cases} \quad (3)$$

以上のように $R_i^T(I, a)$ と戦略 $\sigma_i^T(I)(a)$ を更新していき、最終的には式 4 のように情報集合への到達確率で重み付けされた各イテレーションでの戦略の平均を算出する. 2 人ゲームでは Blackwell の定理 [8] よりこれがナッシュ均衡戦略に収束されることが保証されている.

$$\bar{\sigma}_i^T(I)(a) = \frac{\sum_{t=1}^T \pi_i^{\sigma^t}(I)(a)}{\sum_{t=1}^T \pi_i^{\sigma^t}(I)} \quad (4)$$

CFR は 3 人以上の多人数ゲームでも実装上はそのまま適用が可能である. ナッシュ均衡戦略への収束性は理論的にはなされていないがコンペティションの成績などから有効性が報告されている [6]. しかしながら計算に 3 人ゲームでさえ 16GB や 64GB の RAM を用いて数週間から数ヶ月を要することや, 4 人以上のテキサスホールデムでは空間計算量的に 2 人ゲームのように直接 CFR を適用できないことが多人数ポーカーでの CFR の問題点となっている.

3.3 ポーカーでの情報集約や状態抽象化

ϵ -ナッシュ均衡戦略を生成するにあたり, 実際のゲームを直接扱うことは空間計算量的に困難なので, 手札などについての集約・抽象化を行う必要がある. 概して抽象化の度合いが低いほど本来のゲームに近い戦略が生成できるため望ましいといえる.

3.3.1 手札に関する抽象化

スートについてはテキサスホールデムではフラッシュかどうか判断できればよいので, 4 種のスートの組み合わせをそのまま扱うのではなく, 同種かどうかだけを判断することで状態数が削減される. しかし ϵ -ナッシュ均衡戦略を計算する上ではスートの抽象化だけでは不十分であり, より大きな集約を行う必要がある. より大きく手札を集約する手法として, カードの強さやポテンシャルといった基準を設けて集約する Bucketing という手法が用いられてい

る. Bucketing の基準としてはカードの勝率や, 手札の最終的な勝率の 2 乗の平均が用いられている [9].

3.3.2 選択肢に関する抽象化

今回扱っていないレイズ額に制限を儲けない No Limit ルールのポーカーでは適したベット額を推定する Action Abstraction [10] といった選択肢の抽象化・集約手法もある.

3.3.3 部分木のみを考慮することによる抽象化の緩和

特定の部分木では専用に計算した低抽象な戦略を適用するという, Strategy Stitching [11] という手法も提案されている. これは例えば初期局面からコール・コール・レイズと進んだ局面 (ccr) からの部分木のみで CFR を実行することで, より少ないノードだけを考慮すれば良いことになり, 手札の抽象度を下げる事が可能になり, 結果としてよりよい戦略を生成できるというものである.

4. 提案手法

参加人数に対して「融通がきく」ゲームでは, ゲームへの参加人数が少し増減してもゲームの性質はそれほどは変わらないことが予想される. そこで CFR などの ϵ -ナッシュ均衡戦略の計算が難しくなってくるであろう大きな人数のゲームに対し, それよりも少ない人数での戦略を適用することを提案する.

4.1 相手プレイヤーの行動の除外と少人数のゲームでの ϵ -ナッシュ均衡戦略の適用

特定プレイヤーの行動を除外することでゲーム状態の抽象化を行う. これにより例えば 4 人ゲームを 3 人ゲームとして考えられるようになり, 3 人ゲームでの戦略を適用できるようになる. プレイヤ行動の除外については今回はゲーム木上での現在状態までの行動のパス (行動シーケンス) の変換によって行う.

基本的には特定のプレイヤーの行動をスキップ・省略して行動シーケンスを再構成することで, 実際のゲームより少人数のゲームへの変換を行う. 直接少人数のゲームでの戦略を適用するために, 想定する少人数ゲームの状況と実際のゲーム上での行動に「実際のゲームではもうレイズが出来ないのにレイズを行う」といった矛盾が起きないように変換を行う必要があると考えられる. そのため, このような矛盾を極力起こさないための変換の方針を 2 つ提案する.

- (1) 全員を一度に除外せず, 1 人ずつ除外していく
- (2) フォールドしたプレイヤーを優先的に除外する

このような方針を立てることで変換前後での選択肢の矛盾や非合法な状態が生じないように処理することが可能となっている.

方針 (1) のための基本的な 1 プレイヤ削減アルゴリズムは Algorithm1 の通りである. 削減する特定のプレイヤーの行動はスキップするが, レイズに関する行動シーケンスの構造や, ベットやベットラウンドの回数は極力保持するよ

Algorithm 1 シーケンスの変換を用いた1プレイヤー削減

```
//変換前の行動シーケンスを  $s[n]$ 
//変換後の行動シーケンスを  $s\_new[]$  とする
 $j \leftarrow 0$ 
// $s[n]$  を前から読み込んでいく
for  $i = 0$  to  $n - 1$  do
  if ベットラウンドが  $s\_new[j] > s[i]$  then
    //局面  $s[i]$  を次のラウンドまで進める
  else if 局面  $s[i]$  の手番プレイヤー=除外対象 then
    //除外対象プレイヤーの行動はレイズ以外反映しない
    if  $s[i]$ =レイズ then
       $next\_raise \leftarrow true$ 
    end if
  else
    //その他のプレイヤーの行動は、仮想ゲームの非合法手や
    //レイズのフラグが立ってない限りそのまま
    if  $s[i]$ =コール  $\wedge$   $next\_raise=true$  then
       $s\_new[j] \leftarrow$  レイズ
       $next\_raise \leftarrow false$ 
    else if  $s[i]$  が  $s\_new[j]$  では非合法手 then
       $s\_new[j] \leftarrow$  コール
    else
       $s\_new[j] \leftarrow s[i]$ 
    end if
     $j \leftarrow j + 1$ 
  end if
end for

//変換前後のベットラウンドを揃える処理
while ベットラウンドが  $s\_new[j] > s[i]$  do
   $s\_new[j]$  を一手戻す
   $j \leftarrow j - 1$ 
end while

//変換前後の現在のラウンドのレイズ数を揃える
while 現ラウンドのレイズ数について  $s\_new[j] < s[i]$  do
   $s\_new[j] \leftarrow$  レイズ
   $j \leftarrow j + 1$ 
end while

return  $s\_new[]$ 
```

うにしている。実際に想定する少人数のゲームでの戦略を適用するためには、Algorithm1 を想定する人数まで繰り返し実行することになる。

行動シーケンスの変換の具体例については図5に示す。ここでは4人ゲームを合法的な2人ゲームになるまで変換を繰り返し行なっている。

5. 実験

5.1 実験設定

実験はルダックホールデムを6人まで同時に行える多人数ゲームに拡張したもので行う。用いるカードの種類と枚数は最大6人で行うことを想定し、7種類・各2枚の合計14枚とした。それ以外の点については従来のルダックホールデムのルールと共通である。

ゲームは3人ゲームから6人ゲームまでそれぞれ行う。



図5 提案手法の適用例

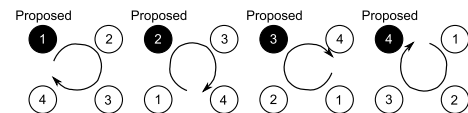


図6 プレイ順序

n 人のゲームでのプレイヤーの組み合わせは、取りうるプレイヤーの組み合わせが多いため今回は提案手法を取るプレイヤーが1人、残りの $n - 1$ 人は n 人ゲームでの CFR で生成した ϵ -ナッシュ均衡戦略を取るプレイヤーとする。提案手法のプレイヤーは前述の行動シーケンスの変換を用いて、少人数のゲームへ落とし込み、2人ゲームから $n - 1$ 人ゲームまでの CFR で生成した ϵ -Nash 均衡戦略を適用する。抽象化するために削減するプレイヤーの選択については前述の方針の通り、フォールドしているプレイヤーから選択していき、想定人数まで達しなかった場合は自分以外のプレイヤーからランダムに選択するものとする。

対戦については、事前に各プレイヤーの手札をランダムに20万ゲーム分生成し、プレイヤーの行動順序の組み合わせ毎に同じ手札の組を用いてゲームを行う。手法の比較は最終的な提案手法を用いたプレイヤーの平均報酬を用いる。

提案手法のプレイヤー以外のプレイヤーは同一の戦略を取るため、行動順序の組み合わせは n 人のゲームなら図6のように提案手法プレイヤーの順序を変えた n 通りについて行う^{*5}。

提案手法、比較手法のプレイヤー共に用いる戦略(確率表)は、CFRによって生成された共通の確率表を利用するものとする。ルダックホールデムはCFRで抽象化せず扱えるため、今回は手札等に関する抽象化は行っていない。CFRの計算時間は各人数毎に24時間^{*6}で固定とし、対戦

^{*5} 5人ゲームなら100万回のゲームを行う。

^{*6} Intel Core i3 530 2.93GHz 1 thread にて

表 2 同一の時間 (24 時間) 実行した場合の
プレイヤー人数ごとの CFR のイテレーション回数

人数	回数
2	212146986
3	14782756
4	1193910
5	127787
6	14854

実験では計算途中の戦略表を適宜出力し^{*7}、それらの組み合わせについて対戦を行った。

5.1.1 CFR の計算時間

24 時間 CFR を実行した場合のイテレーション回数とプレイヤー人数の関係を表 2 に示す。プレイヤーの人数が 1 人増えるたびにイテレーション回数がおおよそ 1/10 になっていることがわかる。これはゲームのノード数が人数が増えるたびに約 10 倍になっているためであると考えられる。

CFR の実装にはイテレーション回数を増やすため、各ノードへの到達確率を用いて Regret の更新を打ち切ったり、モンテカルロシミュレーションを行なって更新していく手法 [12] があるが、多人数ゲームでは更新されるノードがスパースになると考えたため今回は行っていない。

5.2 実験結果

以下実験結果について示す。なお全てのグラフの縦軸は提案手法の平均報酬であり^{*8}、高いほうが良い結果となる。

5.2.1 提案手法の計算時間を固定した場合

提案手法の CFR の実行時間をほぼ 24 時間のところに固定し、比較プレイヤーの CFR の実行時間を提案手法と同等の時間になるまで変化させた場合の対戦結果について図 7 に示す。右に行けば行くほど比較プレイヤーの CFR の実行時間は長くなり、右端の点がほぼ 24 時間 CFR を実行した比較プレイヤーに対する提案手法の平均報酬となる。

5.2.2 対戦相手の計算時間を固定した場合

対戦相手の CFR の実行時間をほぼ 24 時間で固定し、提案手法の CFR の実行時間を比較手法の CFR 実行時間である 24 時間まで変化させた時の対戦結果を図 8 に示す。

例えば図 8(d) では 3 人 CFR や 4 人 CFR を利用した提案手法プレイヤーは比較プレイヤーの 6 人 CFR の実行時間の 10% も実行せずにより多くの報酬を得ていることがわかる。

6. 考察

6.1 人数の変化に対する ϵ -ナッシュ均衡戦略の有効性

今回実験を行った CFR のイテレーション回数では、図 7(a) の 3 人ゲームでこそ若干負け越しているが 4 人以上の n 人ゲームでは図 7(b) から 7(d) までより 1 人減らした $n-1$ 人の CFR を利用したプレイヤーを用いても平均報酬は

^{*7} 例えば 6 人ゲームではイテレーション 1000 回、2000 回、3000 回... という具合にきりの良い回数で確率表を出力した。

^{*8} 単位はチップ枚数、参加費・報酬は 2.2 節を参照

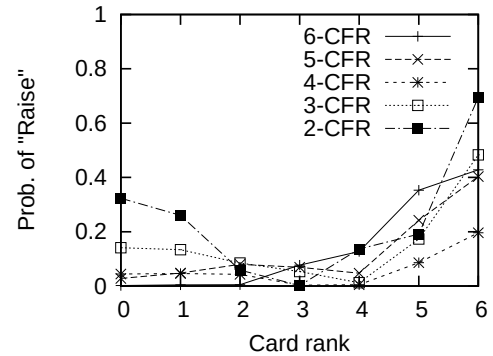


図 9 初期局面でレイズを行う確率
rank はペアでなければ 0 が弱く 6 が強い

ほとんど ± 0 かそれ以上得られていることがわかる。

2 人 CFR を用いた提案手法プレイヤーの戦略は図 8 を見る限り十分に収束しているが、最も近いゲームである 3 人ゲームでも若干負け越しており、それ以上のゲームでは大きく負け越している。

3 人 CFR を用いた提案手法の対戦結果に着目してみると、図 7(b), 7(c) では対戦相手は戦略が収束してきているにもかかわらず報酬は 0 付近で安定し、図 7(d) のプレイヤーが想定より 3 人も多い 6 人ゲームでも対戦相手の戦略が十分に収束していないこともあってか、大きく勝ち越していることがわかる。

また、図 8(d) より若干収束していないと思われる 5 人 CFR (proposed 5-CFR) を利用した提案手法でも 6 人ゲームで図 8(d) 右端や図 7(d) のように 4 人 CFR と遜色ない成績を収めていることから、概して想定人数が元の人数に近いほうが良い結果をあげられるのではないかと考えられる。

6.2 CFR によって生成された戦略の人数による違い

今回の提案手法や比較のプレイヤーの両方で用いられている、CFR で生成された確率表から提案手法のシーケンスの変換の影響を受けない、初期局面でのレイズを行うかの確率を図 9 に示す。

これを見ると人数が少ない時、特に 2 人ゲームでの CFR によって得られる戦略は、特に 4 人以上の CFR の戦略とは異なり、弱いカードであってもレイズをする確率が比較的高く、人数が多くなるとカードが強い時しかレイズしていないことがわかる。人数が少なくなればなるほど明確に勝つための相手がわかるためブラフのような行動が行い易いためではないかと考えられる。逆に人数が多ければブラフを行ったとしても実際に強い手札のプレイヤーが場にいる確率が高いため、ブラフを行うメリットがなくなってしまうのでは無いかと考えられる。

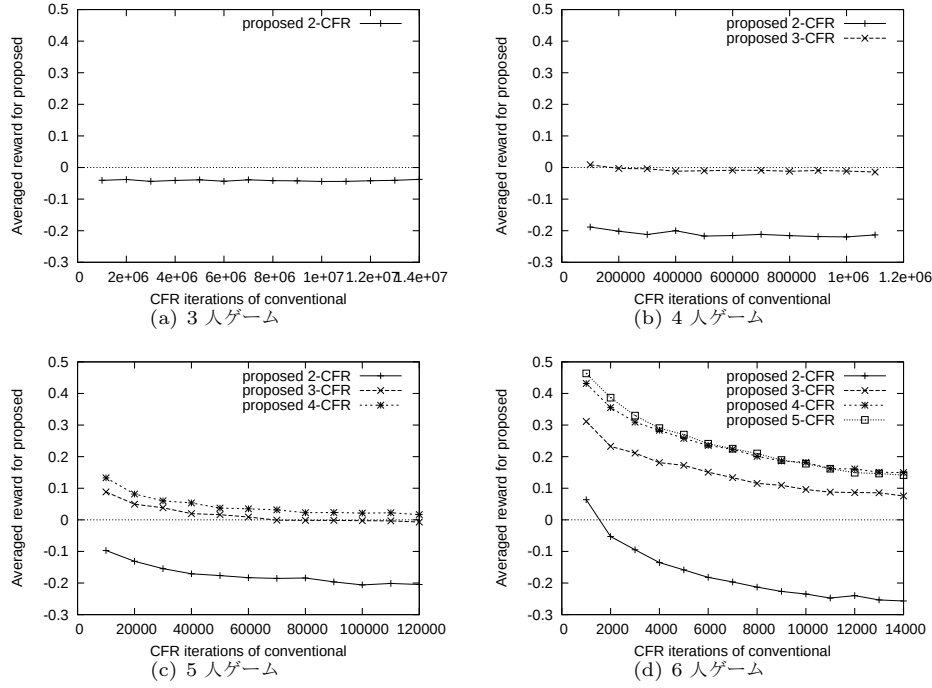


図 7 実験結果 (提案手法の CFR 実行時間を 24 時間で固定)

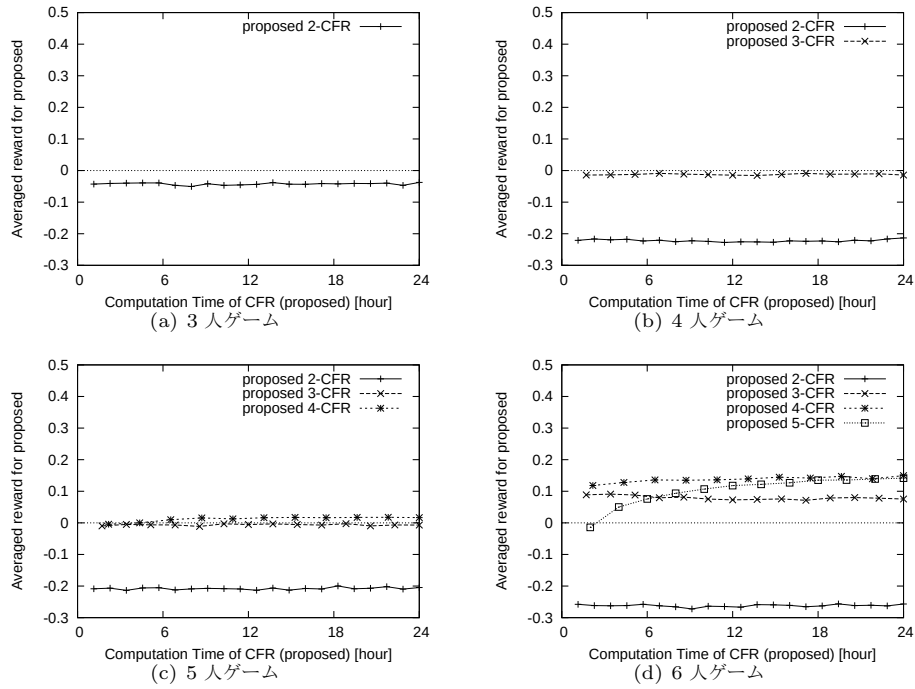


図 8 実験結果 (比較プレイヤーの CFR 実行時間を 24 時間で固定)

6.3 比較プレイヤーの報酬のプレイ順序による依存性

プレイヤーの順序による報酬の違いについて表 3 に示す。この表の対戦結果はお互い 24 時間 CFR を実行した場合の対戦結果であり、実験結果のグラフの右端の点に対応している。表を見て分かる通り、比較プレイヤーは順序によって報酬が大きく変わっていることがわかる。特に提案手法プレイヤー (1) の直後のプレイヤー (2) についてはすべての場合で提案手法のプレイヤーとは報酬の正負が反転しており、提案手法のプレイヤーの報酬が大きかった場合にはそのプレイ

ヤは負け越していることがわかる。これは直後のプレイヤーほど提案プレイヤーの行動に直接影響されやすく、順序が遠くなれば間のプレイヤーの行動が挟まることによって、影響が小さくなるためではないかと考えられる。

6.4 CFR の収束性について

3 人以上の多人数ゲームでの CFR では戦略のナッシュ均衡への収束性については理論的には保証はされていない。しかし図 7, 8 のようなグラフの結果から人数が増えてもイ

表 3 プレイ順序による報酬の違い

対戦組合せは例えば“2 vs 3”は提案手法は2人CFRを利用し、比較プレイヤは3人CFRを用いて3人ゲームを行なっている。プレイヤの番号はプレイの順序・隣接関係を表している。平均報酬の小数点以下第4位以下は切り捨てている。

対戦 組合せ	提案 1	比較プレイヤ				
		2	3	4	5	6
2 vs 3 (3 のみ)	-0.037 -0.002	0.083 0.003	-0.045 -0.000			
2 vs 4 3 vs 4 (4 のみ)	-0.213 -0.014 -0.001	0.057 0.010 0.000	0.076 0.000 0.000	0.079 0.003 0.000		
2 vs 5 3 vs 5 4 vs 5 (5 のみ)	-0.204 -0.007 0.016 -0.001	0.130 0.005 -0.006 0.002	0.152 0.006 -0.010 -0.001	0.153 0.001 -0.015 -0.001	-0.232 -0.006 0.014 0.001	
2 vs 6 3 vs 6 4 vs 6 5 vs 6 (6 のみ)	-0.256 0.075 0.149 0.141 0.000	0.086 -0.020 -0.039 -0.040 -0.002	-0.064 0.011 0.016 0.018 0.002	0.085 -0.013 -0.047 -0.037 0.003	0.068 -0.022 -0.035 -0.038 -0.004	0.080 -0.030 -0.043 -0.042 0.001

テレセッション回数を増やすほど報酬が得られるような方向へ変化し、十分回数を重ねている部分では報酬が収束していると考えられる。この収束点がナッシュ均衡とどのような関係にあるのかは今後の課題となる。

7. おわりに

本稿ではナッシュ均衡戦略の計算がより困難になってくる、3人以上の多人数ゲームで、相手行動の除外・抽象化を用いて、少人数でのゲームで計算した ϵ -ナッシュ均衡戦略を適用することを提案した。トイゲームを用いた実験であるが、従来では扱われて来なかった人数に対する多人数ゲームにてCFRを適用し、直接そのゲームで実行し十分に学習しきれていない戦略を用いるよりも、小さなゲームで十分に学習した戦略を用いた方が大きな報酬が得られることがわかった。今回の実験の結果は3人テキサスホールデムでのCFRによって得られた戦略が4人以上のゲームでも有効に働く可能性を示唆している。

一方で比較手法のCFRの実行時間を増やしていけば提案手法の報酬が下がっていったり、3人ゲームでは2人ゲームの戦略を適用しても良い結果が得られなかったりしたように、いかなる場合でも人数を減らすことが有効というわけではないこともわかった。今回提案した手法は特に人数が多すぎて、計算時間的に学習が収束しない場合や、状態数が増加してメモリ上で扱えない場合に有効であると考えている。

今後の課題としてはポーカー以外のゲームへの適用可能性が考えられる。相手プレイヤの抽象化手法として行動シーケンスの変換を提案したが、これはプレイヤ間に依存関係の生じる行動の除外に対応できなければ上手いかわない。一方で参加人数が可変なゲームは行動に対して生じる

依存関係は、それほど複雑にはならないとも思われる。そのため今回の提案手法でも行なっている、プレイヤ間に依存性が生じる行動を取るプレイヤの行動を他のプレイヤに引き継がせるなどすることができれば、本手法のような変換は適用可能であると考えられる。

参考文献

- [1] Nash, J.: Equilibrium points in n-person games, *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 36, No. 1, pp. 48–49 (1950).
- [2] Koller, D., Megiddo, N. and Von Stengel, B.: Fast algorithms for finding randomized strategies in game trees, *Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*, ACM, pp. 750–759 (1994).
- [3] Zinkevich, M., Johanson, M., Bowling, M. and Piccione, C.: Regret minimization in games with incomplete information, *Advances in Neural Information Processing Systems*, Vol. 20, pp. 1729–1736 (2008).
- [4] Bowling, M., Risk, N., Bard, N., Billings, D., Burch, N., Davidson, J., Hawkin, J., Holte, R., Johanson, M., Kan, M. et al.: A demonstration of the Polaris poker system, *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*, International Foundation for Autonomous Agents and Multiagent Systems, pp. 1391–1392 (2009).
- [5] Schaeffer, J., Billings, D., Peña, L. and Szafron, D.: Learning to play strong poker, *The International Conference on Machine Learning Workshop on Game Playing* (1999).
- [6] Risk, N. and Szafron, D.: Using counterfactual regret minimization to create competitive multiplayer poker agents, *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*, International Foundation for Autonomous Agents and Multiagent Systems, pp. 159–166 (2010).
- [7] Southey, F., Bowling, M., Larson, B., Piccione, C., Burch, N., Billings, D. and Rayner, C.: Bayes’ bluff: Opponent modelling in poker, *Proceedings of the 21st Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, Citeseer, pp. 550–558 (2005).
- [8] Blackwell, D.: An analog of the minimax theorem for vector payoffs, *Pacific Journal of Mathematics*, Vol. 6, No. 1, pp. 1–8 (1956).
- [9] Johanson, M.: Name of Author: Michael Bradley Johanson Title of Thesis: Robust Strategies and Counter-Strategies: Building a Champion Level Computer Poker Player Degree: Master of Science Year this Degree Granted: 2007.
- [10] Hawkin, J., Holte, R. and Szafron, D.: Automated action abstraction of imperfect information extensive-form games, *National Conference on Artificial Intelligence (AAAI)* (2011).
- [11] Waugh, K., Bard, N. and Bowling, M.: Strategy grafting in extensive games, *Advances in Neural Information Processing Systems*, Vol. 22, pp. 2026–2034 (2009).
- [12] Lanctot, M., Waugh, K., Zinkevich, M. and Bowling, M.: Monte Carlo sampling for regret minimization in extensive games, *Advances in Neural Information Processing Systems*, Vol. 22, pp. 1078–1086 (2009).