不確定不完全情報展開型多人数ゲームにおける相手モデル化による搾取相手の選択

古居敬大[†1]　三輪誠[†2]　近山隆[†1]

ナッシュ均衡的な戦略は多人数ゲームでは有効な戦略であるが，非合理的なプレイヤが存在する場合には必ずしも最適な行動であるとは言えない．本稿では多人数ポーカーゲームにおいて，より搾取が可能であると予想されるナッシュ均衡戦略を取っていないプレイヤを判別し，そのプレイヤのみに応じた戦略を取るプレイヤについての提案する．実験を行ったところ，特定の単純な行動を取るプレイヤに対しては大きく搾取することができ，結果としてナッシュ均衡的な戦略をとったプレイヤより報酬が大きくなる場合があることを確認した．

Opponent Exploitation by Opponent Modeling for Probabilistic Imperfect Information Extensive Multi-player games

KEITA FURUI,[†1] MAKOTO MIWA[†2] and TAKASHI CHIKAYAMA[†1]

A Nash-equilibrium strategy is known to be effective for multi-player games, but this is not always the best strategy because of the existence of naive players. In this paper, we propose a game playing strategy for multi-player poker games. In this strategy, the player detects an opponent who dose not adopt the Nash-equilibrium strategy and exploits the opponent without considering the other opponents. Experimental result show the player with the proposed strategy was able to exploit a naive opponent and get more rewards than the ε-Nash equilibrium player in some cases.

1. はじめに

ポーカーはチェスや将棋のような確定完全情報ゲームとは異なり，カードのシャッフルという不確定な要素がある点や，相手の手札が見られないという不完全情報性があるため，自分の有利不利の判断が困難な状況下で行動を決定しなければならないという課題がある．

ゲーム理論の解のひとつであるナッシュ均衡解[1]的な戦略をとることは，報酬を最大化するための行動としては有効であると考えられる．標準型ゲームとしての記述が困難な展開型ゲームのポーカーについても，ナッシュ均衡戦略を近似するアルゴリズムが提案されてきている[2]．しかし必ずしも相手に合わせた行動が取れないため，明らかに弱いと考えられるプレイヤとの対戦でも大きく勝ち越そうとはしない．そのような観点から相手のモデル化についての研究もまたなされてきている[3)–7)]．

ナッシュ均衡的な戦略をとることは多人数ゲームでは必ずしも最善な行動にはならない．これはナッシュ均衡がプレイヤの中に搾取されやすい行動を取るプレイヤや，利他的な行動を取るプレイヤの存在を考慮しておらず，そのようなプレイヤの行動を利用できるプレイヤの報酬が最大となることがあるからである．

本稿ではある種の多人数ポーカーゲームにおいて，ナッシュ均衡的な戦略を取るプレイヤと単純な行動を取るプレイヤの両方が存在する状況で，より搾取が可能であると予想される単純な行動を取るプレイヤを判別し，ナッシュ均衡的な戦略をとるプレイヤの行動を無視し，単純な行動を取るプレイヤにのみに応じた戦略を取る手法について提案する．実験において単純な行動を取るプレイヤによっては，ナッシュ均衡的な戦略と同程度の報酬を得られることがわかった．

2. ポーカー

不確定不完全情報ゲームであるポーカーには多くの種類が存在する．今回はホールデムポーカーの一種である Texas Hold'em の簡易版である Leduc Hold'em について取り扱う．この種類のポーカーは 1 ゲーム中に

†1 東京大学大学院工学系研究科
Graduate School of Engineering, The University of Tokyo
{furui,chikayama}@logos.ic.i.u-tokyo.ac.jp
†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
makoto.miwa@manchester.ac.uk

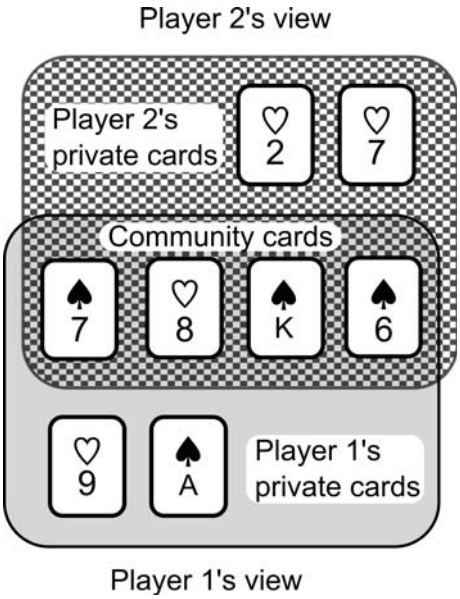

図 1 Texas Hold'em のプレイヤごとに得られる情報

複数のベットラウンドが存在し，ラウンドごとに各プレイヤは（1）賭けを続行するかどうか，(2）賭けるチップを上乗せするかどうか，の選択を行う．場には各プレイヤが共通して用いるコミュニティカードがあり，ラウンドを経るにつれてそのカード数は多くなるといった特徴がある．

2.1 Texas Hold'em

Texas Hold'em (テキサス・ホールデム) は毎年世界選手権も開催されている，最も一般的なホールデムポーカーの 1 つである．

2.1.1 ル ー ル

各プレイヤごとに固有のプライベートカード 2 枚と，5 枚のプレイヤが共有するコミュニティカードがあり，最終的にはこの 7 枚の中から 5 枚を組み合わせてその役の優劣で勝敗を決める．

図 1 にあるとおり各プレイヤは相手のプライベートカードを観測することができないため，相手の行動から相手の強さを推測する必要がある．

勝敗を決めるショーダウンまでには 4 つのベットラウンドがあり，各ラウンドでは賭額を上げたり，途中でゲームから降りることができる．各ラウンドで可能な行動はレイズ（Raise），コール（Call），フォールド（Fold）の 3 つであり，

- レイズ : 図 2 のように最も賭けているプレイヤのチップ枚数から更にチップを上乗せしてゲームを続行する．リミットルールでは上乗せできるチップはラウンドごとに決まっていて，また，各ラウンドで行えるレイズ回数に制限がある．
- コール : 図 3 のように賭けるチップを最も賭けているプレイヤのチップ枚数に揃え，ゲームを続行する．
- フォールド : 賭けを中止してゲームから降りる．賭けたチップはそのプレイヤに返還されない．また，ゲームから降りたプレイヤのプライベートカードは公開されない．

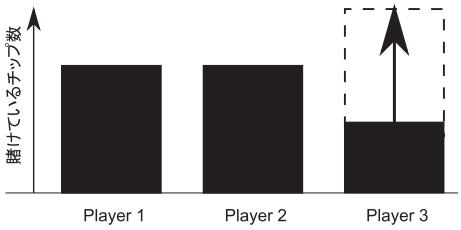

図 2 レイズ

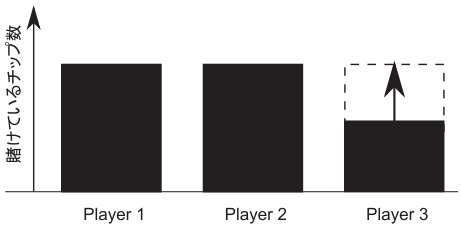

図 3 コール

ゲームを続行するプレイヤ間で賭けるチップ枚数に合意がなされれば次のラウンドに移る．各ラウンドでは以下のようにカードが配布，公開される．

(1) preflop:プライベートカードを各プレイヤに 2 枚ずつ配布する
(2) flop:コミュニティカードが 3 枚公開される
(3) turn:コミュニティカードが更に 1 枚公開
(4) river:最後のコミュニティカードが公開
(5) (ショーダウン):互いのプライベートカードを公開し，役の優劣で勝敗を決める

賭けたチップは最後までゲームから降りなかったプレイヤの中で，ショーダウン時に最も強い役を持つプレイヤに報酬として支払われる．また，1 人以外のプレイヤが全員フォールドした場合はその場でゲームは終了し，そのプレイヤに賭けたチップが報酬として与えられる．

賭けから降りた場合はそれまでに賭けた金は戻ってこないため，勝てそうにないときは適度なタイミングでゲームを降りるなどの判断が求められる．一方でブラフのような強気なプレイによって，相手をフォールドさせるのもまた有効な戦略である．

ルールによっては参加費（ante）やブラインドベッ

トという初めに払う賭金がある．テキサスホールデムでは 1 番目のプレイヤが 5 枚，2 番目のプレイヤが 10 枚のチップをブラインドベットとして賭けるのが一般的である．

2.2 Leduc Hold'em

Leduc Hold'em[6]（ルダック・ホールデム）は Texas Hold'em での戦略的な要素をできるだけ保持しつつゲームの規模を縮小させたもので，主にホールデムポーカーゲーム自体の分析やゲームプレイヤ研究の実験環境として用いられている．

テキサスホールデムからの変更点や主な特徴は，

- カードの枚数は柄 2 種類，数字 3 種類の計 6 枚（3 人ゲームなら数字 4 つの 8 枚）と少数
- 役はペアと数字の大小のみ
- ラウンドは 2 つで，各ラウンドでレイズ回数は 2 回まで
  (1)　preflop : プライベートカード 1 枚ずつ配布
  (2)　flop : コミュニティカード 1 枚公開
- ベット回数はラウンドごとに 2 回まで
- レイズ額は preflop で 2 枚，flop で 4 枚
- 全員が参加費としてチップ 1 枚をゲーム開始時に賭ける

というものである．

2 人ゲームの場合，行動に関するゲーム木は図 4 のようになる． ゲームは図 4 の上部のルートノードから

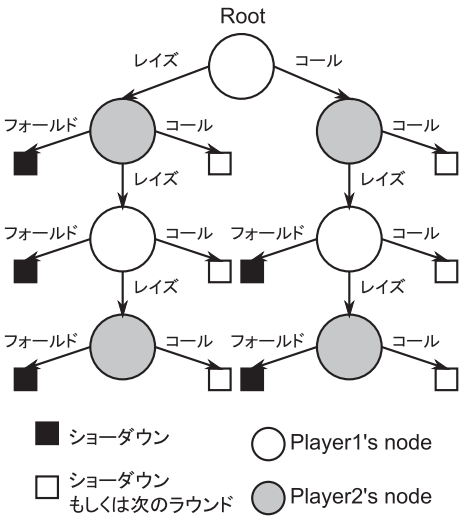

図 4　2 人 Leduc Hold'em

ら始まる．1 番目のプレイヤはルートノードでは賭け額が相手と等しいためフォールドする理由はなく，そのため基本的にはレイズとコールの 2 択となる．次にルートからコールした局面でも同様のことが言える．それ以外のノードでの行動はレイズとコール，フォールドの 3 択となる．図 4 の■ノードに到達したときは片方のプレイヤがフォールドしたとして，ゲームは終了となる．□ノードについてはラウンドが preflop であれば，次の flop ラウンドが開始され，またルートノードからの行動選択が行われる．flop ラウンドであれば□ノードに到達した時点でショーダウンとなり，カードの役の優劣で勝敗を決定する．

3. 関 連 研 究

3.1 ナッシュ均衡

ナッシュ均衡解はゲーム理論における，特に非協力ゲームの解のひとつである．プレイヤは全員が非協力的で合理的であるという仮定のもと，どのプレイヤも他プレイヤの行動を固定した場合，現在の行動から外れると報酬が下がってしまう，そのような均衡がナッシュ均衡である．有名なナッシュ均衡の例としては表 1 の囚人のジレンマがある．

	B: 黙秘	B: 自白
A: 黙秘 (協力)	(1, 1)	(4, 0)
A: 自白 (裏切り)	(0, 4)	(3, 3)

表 1　囚人のジレンマ

ナッシュ均衡解は 2 人零和確定完全情報ゲームでは最適解にあたる．また不確定不完全情報ゲームにおけるナッシュ均衡は各行動を確率的に分配した混合戦略の形を取ることが知られている．

ナッシュ均衡戦略は相手全員が完全に合理的な行動を取ることを仮定しており，行動は守勢となり，必要以上のリスクは負わない傾向にある．このため，相手が仮に弱いと考えられるプレイヤであったとしてもその相手から必要以上には攻めようとはしないという特徴がある．このように全員が合理的だという仮定が十分に満たされない場合は必ずしもナッシュ均衡戦略は最適な行動になるとは言えない．

じゃんけんゲームを例にとって考える．じゃんけんゲームのナッシュ均衡はすべての手を等しく 1/3 ずつ出すことである．しかしながらそのような戦略ではどんな相手に対しても，平均的に負け越すことはないが，勝ち越すこともまた出来ない．ここでさらに 3 人でじゃんけんを行うことを考える．1 人がナッシュ均衡戦略をとっていたとして，もう 1 人が仮にグーしか出さなかった場合，残りの 1 人がパーを出し続けるとそのプレイヤの必勝となることは明らかである．

このように，特に複数人ゲームにおいては，明らかに合理的ではない，弱いプレイヤや利他的なプレイヤ

が存在しているような状態では必ずしも報酬は最大化されない，といえる．

3.1.1 ナッシュ均衡戦略の近似

標準型ゲームとしての記述が可能な場合，ある種の線形計画問題を解くことでナッシュ均衡解を求めることは可能である．しかし厳密なナッシュ均衡解をポーカーのような状態数の比較的大きな[⋆1]展開型ゲームについて求めることは計算量的に実質不可能である．そのような場合には状態の抽象化を行って状態数を減らすことで，ナッシュ均衡の近似戦略（ε ナッシュ均衡戦略）を求めることが戦略として有効であることが知られている．Texas Hold'em についてはラウンドの抽象化やカードの強さの抽象化（Bucketing）を行う．概してナッシュ均衡の近似戦略は元のゲームに近ければ近いほどよく，そのため抽象度が小さいほど戦略として優れているといえる．

3.1.2 CounterFactual Regret minimization

CounterFactual Regret minimization[2]（CFR）は 2 人展開型ゲームにおける，期待損失の最小化によってナッシュ均衡戦略に収束する反復アルゴリズムである．不完全情報の組み合わせの多さから空間計算量的に標準形で扱うのが困難なポーカーゲームについて，不完全情報をひとつの情報セットとして集約することで計算機が保持する状態数を削減し，より少ない抽象度でナッシュ均衡戦略の近似を行うことが可能となっている．

2 人ゲームにおいて CFR は繰り返し計算を行うことで，ナッシュ均衡戦略への収束が保証されている．しかし 3 人以上の多人数ゲームではその理論的保証がないが，アルゴリズム自体は多人数ゲームについても適用可能であり，それによって生成された戦略を取るプレイヤの強さもまた示されている[8]．

3.2 相手モデル化やその最適反応

相手モデル化（Opponent Modeling）や最適反応（Best Response）は相手の戦略をモデル化し，それに対して最適な行動を取るという考え方である．最適反応 s_i^* は，自分の戦略を s_i，自分以外の戦略を s_{-i}，自分の報酬を u_i として，

$$u_i(s_i^*, s_{-i}) \geq u_i(s_i, s_{-i}) \quad (1)$$

で定義される[9]．

相手のモデル化については様々な研究がなされており，ここでは大きく次の 3 つを紹介する．

ひとつに，ゲーム中に相手の行動を集計してヒストグラムを作成することで，相手のモデルを作成しゲーム木探索での行動を予測するもの[4] がある．これは頻出する場面でのモデル化は正確に行えるが，あまり発生しない局面での正確な行動モデルの生成が遅くなると考えられる．

次に，予め用意していた相手モデルに当てはめて戦略を調整するもの[7] がある．これは相手行動に対応できるようになるまでの時間は，モデルを生成するより短くて済むが，相手の戦略推定の精度は劣ると考えられる．

また予め対戦する相手を想定して，それに応じた行動を取るように訓練を行うもの[3),5)] もある．

4. 提案手法とその実装

多人数ゲームにおいては，特定の相手の行動に着目して搾取的な行動をとることで，それ以外のプレイヤの行動についてはあまり考慮しなくても，最終的な報酬は高くなる事は考えられる．

今回は，ナッシュ均衡的な戦略を取るプレイヤと，単純な行動を取る相手プレイヤとが存在する多人数ゲームを想定する．そこでよりナッシュ均衡的でない戦略を取っているプレイヤのほうが搾取が可能であるという仮定を置き，そのプレイヤに対して搾取的な行動を取ることを考える．

本節では，そのような行動を取るための具体的な方法の 1 つとして，

- 相手のモデル化
- 特定の相手のみに着目した搾取的な戦略
- 搾取対象の選択

について提案し説明する．1 ゲーム中における提案手法全体の流れとしては，

(1) 相手の行動履歴からナッシュ均衡戦略をよりとっていないと考えられるプレイヤを選択
(2) 行動ヒストグラムから選択したプレイヤについてのモデルを作成
(3) 1 ゲームが終了するまで以下を繰り返す
  (a) 相手モデルと行動シーケンスに基づいた仮想的な 2 人ゲーム木の探索
  (b) 選択したプレイヤがフォールドした場合は第 3 のプレイヤの行動についてモデル化し探索
(4) 行動ヒストグラムの更新

となる．

今回の提案手法の実装については 3 人 Leduc Hold'em に向けて行っている．

⋆1 経済学分野で取り上げられるような標準型ゲームの状態数は囚人のジレンマの例のような $2 \times 2 = 4$ など非常に小さく，それと比較して大きな，と形容している

4.1 相手のモデル化

4.1.1 モデル化の前提

今回は相手のモデル化を行うにあたり以下のように仮定する．

(1) プレイヤの行動は他プレイヤの行動に依存しない

(2) 自らのカードと場のカードのみによって行動を決定する

これらのプレイヤの行動の仮定は現実的であるとは言えないが，今回は特に単純なプレイヤへの搾取を念頭に置いているためこのような仮定でモデル化を行った．

4.1.2 相手行動のヒストグラムの作成

相手の特定の状況における行動をヒストグラムとして記録し，相手の混合戦略をモデル化する．

戦略モデルは

- ラウンド（preflop, flop）
- 場のカードの種類
- レイズ数（フォールド，レイズが可能かどうか）
- 1 ゲーム終了時に公開される相手のプライベートカード（相手がフォールドしなかった場合）

によって 60 [⋆1] の状態に分別し，その状況下での行動（レイズ，コール，フォールド）の選択回数を 1 ゲームが終了するたびに記録する．ヒストグラムから推定する相手プレイヤの混合戦略は，

$$p(n,a) = \frac{c_{n,a}}{\sum_{a'} c_{n,a'}} \tag{2}$$

とした．ただし c を頻度，n を状況，a を行動とする．このモデル化については，

- 観測していない状況についてのモデル化ができない
- 現れにくい状態についてのヒストグラムが更新されにくい

といった問題がある．そのような観測していない状態については今回は，簡単に完全にランダムな行動を取るようなモデルとして扱った．

また相手がフォールドした場合については相手のプライベートカードを観測できないため，今回は特定のプライベートカードにおけるフォールド回数を，

$$c_{n,a=f} = c_{n',a=f}/4 \tag{3}$$

というように単純に Leduc Hold'em カードの種類数である 4 で割ることで近似している．

⋆1 preflop : カード 4 種類，flop : カード 4 × 4 = 16 種類で 4 + 16 = 20．レイズ数は 0，1，2 の 3 パターンであるから 20 × 3 = 60．今回はカードの種類の抽象化，集約を行なっていない．

図 5 行動シーケンスの変換

4.2 特定の相手のみに着目した搾取的戦略の実行

4.2.1 第三者プレイヤの行動の除外

特定の相手のみの行動に着目するために，第三のプレイヤの行動については原則除外して，仮想的な 2 人ゲーム木上を探索する．除外については以下のような方法で行うものとする．

(1) 行動シーケンスから第三者の行動を除外

(2) 2 人ゲーム木の上で見られないような繰り返し行動がみられる場合，その部分を除外してシーケンスを短縮する．

具体例として図 5 を挙げる．ここでは 3 人ゲームの行動シーケンス/crrcc/crc について 1 番目のプレイヤが 3 番目のプレイヤのみに着目した 2 人ゲームシーケンスへの変換を行っている．

4.2.2 仮想的な 2 人ゲーム木の探索

上記のようにして 3 人ゲームの行動シーケンスを 2 人ゲームのものに変換した．そのシーケンスを用いて仮想的な 2 人ゲーム木上で図 6 のような Expectimax 探索[10] を行う． この探索は自プレイヤの行動ノードについては評価値最大の子ノードを選択し，そのノードでの評価値とする．相手プレイヤのノードではそのノードでの混合戦略とその子ノードの評価値を用いてそのノードの期待値を算出し，それを評価値とする．

レイズの回数の制限については第 3 者の行動の影響を受ける．そのため仮想的な 2 人ゲーム上で必ずしもレイズが取れない場合がある．その場合については今回は単純にコールをすることとした．

4.3 搾取対象の選択

4.3.1 ε ナッシュ均衡戦略を取っている確率

予め算出しておいた ε ナッシュ均衡戦略の確率表を

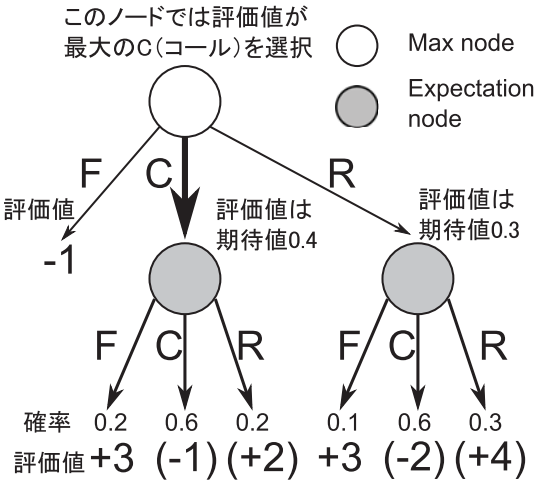

図 6 Expectimax 探索

用いて，ナッシュ均衡的な戦略を取っているかを判定する．具体的には，行動を取る確率の対数に-1 を掛けたものの平均である，

$$-\frac{\sum_{n=0}^{N-1}\log(P_{\varepsilon Nash}(n,a))}{N} \quad (4)$$

が最大のプレイヤをよりナッシュ均衡でない行動を取っていると判断し，そのプレイヤをモデル化と搾取的戦略の実行対象にする．

5. 評　　価

提案手法を取るプレイヤと，ナッシュ均衡的な戦略を取るプレイヤと，単純な行動を取るプレイヤの 3 人で Leduc Hold'em を繰り返し行う．プレイの順序は 3!=6 通りあるので，順序ごとに配布されるカードを固定して対戦を行う．

それぞれのプレイヤは相手の 2 人のプレイヤがそれぞれがどんな戦略を取るのかを知らされておらず，提案手法としてはどちらが単純な行動のプレイヤかを判断する必要がある．

実験で用いたプレイヤとその組み合わせは以下のようになっている．

(1) プレイヤ 1（提案手法とその比較）
- A:提案手法
- B:ナッシュ均衡的戦略
- C:ランダム

(2) プレイヤ 2（ナッシュ均衡戦略）
- D:ナッシュ均衡的戦略 (=B)

(3) プレイヤ 3（単純な行動を取るプレイヤ）
- E:常にレイズ（レイズができないときはコール）
- F:常にコール
- G:ランダム (=C)

ランダムな行動を取るプレイヤ（C，D）とはとりうる選択肢を当確率にランダムに選ぶプレイヤとする．対戦はプレイヤの組み合わせ 3 × 1 × 3 ＝ 9 通りの 1 つにつき，プレイヤの順序（6 通り）ごとに 5000 回ずつ繰り返しゲームを行い，得た報酬の平均を比較する．

カードの組み合わせと混合戦略の行動決定に関わる乱数については，予め生成しておくことで組み合わせ・順序ごとに同一のものを使用した．

5.1 結　　果

対戦結果の平均報酬は表 2 のとおりである．

組	A:提案手法 B: ε ナッシュ C:ランダム	D:ε ナッシュ	E:常にレイズ F:常にコール G:ランダム
ADE	+0.9071667	+0.9268167	-1.833983
ADF	+0.0878	+1.10555	-1.19335
ADG	+0.6865	+1.0448	-1.7313
BDE	+0.7276667	+0.7272667	-1.4549333
BDF	+0.7048833	+0.7054	-1.410283
BDG	+0.7045167	+0.7045	-1.4090167
CDE	-1.8579167	+0.8419333	+1.0159833
CDF	-0.8307667	+1.4180833	-0.5873167
CDG	-0.9936333	+1.98845	-0.9948167

表 2 組み合わせ別の対戦結果

5.2 提案手法について

表 2 のうち上部 3 分の 1（ADE，ADF，ADG）が提案手法を用いた対戦結果となっている．提案手法は少なくとも単純な行動を取るプレイヤ（E，F，G）に対しては負け越すことはなかった．また，常にレイズを行おうとするプレイヤに対して（ADE）は大きく搾取ができており，ε ナッシュ均衡戦略のプレイヤとほぼ同等の報酬が得られた．

5.3 A と B の組の各プレイヤの平均報酬の比較

ADE と BDE を比べると A は B より多くの報酬を得られているが，D の ε ナッシュ均衡戦略を取るプレイヤもまた，ε ナッシュ均衡戦略を取る B より提案手法の A が相手の方がより多くの報酬を得ていることもわかる．これは B よりも A のほうが D に搾取されているためであると考えられる．

単純な行動の相手プレイヤが E 以外の場合についても，基本的に A が相手の時のほうが D の報酬は大きくなっていることがわかる．

5.4 A と C の組の比較

提案手法の代わりにランダムな行動を取るプレイヤを用いている場合（CDE，CDF，CDG）は全体として大きく負け越している．このことから，同じ状況で，

ランダムな行動を取るプレイヤ C と同様の選択の取る可能性のある提案手法 A が，C に比べてより他プレイヤ，特に単純な行動を取るプレイヤから搾取を行うことができていると考えられる．

ε ナッシュ均衡戦略を取っている D と B に関しては安定して高い報酬が得られている．しかし，CDE の組み合わせでは E の常にレイズをとろうとするプレイヤに負け越している．

5.5 A（C）と D の組の比較

提案手法の ADE や，ランダム行動と常にレイズの CDE はナッシュ均衡と平均報酬が同等以上になっているが，これはナッシュ均衡戦略は相手のレイズに対してフォールドしやすい傾向にあり，それらの組み合わせではナッシュ均衡を取っているプレイヤ以外がレイズを多く取っているからであると推察できる．提案手法の場合はレイズを取るプレイヤから多く搾取できる機会がよりあったため，大きく搾取を行うことができたと考えられる．同様に CDE の場合も，ランダムな行動を取るプレイヤ C と常にレイズを取るプレイヤ E がレイズを取ることでナッシュ均衡を取るプレイヤはフォールドしてしまい，その後 C がフォールドすることで結果，高額のチップがショーダウンを経ることなく E に渡ったものだと考えられる．この場合，C は E に対して利他的な行動を取っていると見なせる．ADH と CDE の個々の順序別の結果は表 3，4 のようになっている．

順序	A:提案手法	D:ε ナッシュ	E:常にレイズ
ADE	+0.9996	+0.1908	-1.1904
AED	-0.0524	+1.7513	-1.6989
DAE	+0.6113	+0.9138	-1.5251
EAD	+1.1085	+0.9911	-2.0996
DEA	+1.273	+0.8193	-2.0923
EDA	+1.503	+0.8946	-2.3976
平均	+0.9071667	+0.9268167	-1.8339833

表 3 提案手法の順序別対戦結果（ADE）

順序	C:ランダム	D:ε ナッシュ	E:常にレイズ
CDE	-2.4576	+0.1572	+2.3004
CED	-2.4766	+1.0527	+1.4239
DCE	-2.2622	+0.5152	+1.747
ECD	-1.4875	+1.4954	-0.0079
DEC	-1.3157	+0.9108	+0.4049
EDC	-1.1479	+0.9203	+0.2276
平均	-1.8579167	+0.8419333	+1.0159833

表 4 CDE の順序別対戦結果

ADE について，今回は提案手法の報酬は ε ナッシュ均衡戦略を取っているプレイヤの報酬より若干低くなっているが，別に異なるデータセットについて実験したところ，表 5 のように勝ち越す場合があることも確認している．順序別での平均報酬には表 3 とは大きく差があるが，ADE や EDA の順序では ε ナッシュ均衡戦略のプレイヤより報酬を大きくする傾向が見られ，また全体としての平均は大きく変わらないことも見て取れる．

順序	A:提案手法	D:ε ナッシュ	E:常にレイズ
ADE	+1.7904	-0.2341	-1.5563
AED	+0.4115	+1.4262	-1.8377
DAE	+0.4132	+1.4918	-1.905
EAD	+0.8818	+0.7272	-1.609
DEA	+1.0602	+1.3192	-2.3794
EDA	+1.1865	+0.6871	-1.8736
平均	+0.9572667	+0.9029	-1.8601667

表 5 提案手法の順序別対戦結果
（ADE，別データセット）

5.6 その他，まとめ

提案手法の搾取相手の選択については，ほぼ最初の数回のゲームで単純な行動を取るプレイヤを確定できており，それ以降はそのプレイヤがフォールドするまでは，そのプレイヤについてのモデル化・搾取的戦略をとっていた．今回は相手プレイヤとしてナッシュ均衡的な行動を取るプレイヤと単純な行動を取るプレイヤとの対戦実験を行ったため、それら相手プレイヤの行動の差がはっきりと現れたためであると考えられる。

実際のゲームでは、相手プレイヤとして

- ε ナッシュ均衡戦略を取るプレイヤのみが存在する
- ε ナッシュ均衡戦略を取るようなプレイヤが存在しない

場合も考えられる。報酬を最大化するためには、前者では ε ナッシュ均衡戦略を取るのが無難であり、後者や、今回実験した状況では ε ナッシュ均衡戦略を取るか、特定の相手プレイヤに着目した搾取的な戦略を取るかを選択する必要が生じると考えられる。

今回提案した手法では特定の相手以外の行動をほとんど除外して戦略を決定している．しかしながら一部の結果で提案手法が ε ナッシュ均衡戦略に同等もしくは勝ち越すことができたのは，相手が自プレイヤらの行動に影響を大きく受けて行動を決定している場合であると考えることができる．また提案手法についての搾取されやすさも考える必要がある．このことから完全に特定相手の行動を除外して平均的に勝ち越すことは困難であり，少なくとも相手の行動から何らかの情報や特徴を抽出した上でどの程度相手行動を考慮するかを検討する必要がある．

6. おわりに

本稿では多人数ポーカーゲームにおける，単純な行動を取る搾取が可能なプレイヤが存在している場合について，合理的な行動を取っている第 3 者プレイヤの行動を考慮しなくても，大きな報酬が得られる可能性のある戦略についての提案と実装，評価を行った．

提案手法を用いた場合の対戦結果は，単純な行動を取るプレイヤの行動がナッシュ均衡的なプレイヤの行動に大きく影響を与える場合には，その合理的なプレイヤより報酬が大きくなる場合があることがわかった．

ε ナッシュ均衡戦略を取るプレイヤは安定して報酬を得ており，その報酬を超えるのは，行動を全く考慮しない場合には難しいこともわかった．

多人数ゲームでのゲーム木の大きさは，プレイヤ人数の指数オーダで増えるため，ε ナッシュ均衡的な戦略を計算すること自体がプレイヤ人数が増えるにつれて計算量的に困難になる[⋆1]ことから，より正確な合理的行動を行おうとするより，相手の行動が合理的かを判断して，より非合理的な行動を取るプレイヤを注目していくことがさらに重要になってくると予想できる．

多人数ゲームにおける探索は考慮すべき要素が多く，またゲームの性質に大きく依存する．そのため何をモデル化し，何を抽象化するべきかが 2 人ゲーム以上に大きな課題となると考えられる．

今後の課題としては，

- より正確かつ柔軟な相手行動のモデル化
- 提案手法の Texas Hold'em への適用やプレイヤの人数が増え状態数が増大したときに、どのプレイヤやゲーム状態に注目するかの考察
- どのようなプレイヤが搾取可能か、またそのようなプレイヤの行動要素の抽出方法の検討

が挙げられる．

参 考 文 献

1) J.F. Nash. 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) M. Zinkevich, M. Johanson, M. Bowling, and C.Piccione. Regret minimization in games with incomplete information. *Advances in Neural Information Processing Systems*, Vol.20, pp. 1729–1736, 2008.
3) M.Johanson, M.Zinkevich, and M.Bowling. Computing robust counter-strategies. *Advances in Neural Information Processing Systems*, Vol. 20, pp. 721–728, 2008.
4) D.Billings, A.Davidson, T.Schauenberg, N.Burch, M. Bowling, R. Holte, J. Schaeffer, and D. Szafron. Game-tree search with adaptation in stochastic imperfect-information games. *Computers and Games*, pp. 21–34, 2006.
5) M.Johanson and M.Bowling. Data biased robust counter strategies. In *Twelfth International Conference on Artificial Intelligence and Statistics*, pp. 264–271. Citeseer, 2009.
6) F. Southey, M. Bowling, B. Larson, C. Piccione, N.Burch, D.Billings, and C.Rayner. Bayes ' bluff: Opponent modelling in poker. In *Proceedings of the 21st Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 550–558. Citeseer, 2005.
7) Sam Ganzfried and Tuomas Sandholm. Game theory-based opponent modeling in large imperfect-information games. In *The 10th International Conference on Autonomous Agents and Multiagent Systems - Volume 2*, AAMAS '11, pp. 533–540, Richland, SC, 2011. International Foundation for Autonomous Agents and Multiagent Systems.
8) N.A. Risk and D.Szafron. Using counterfactual regret minimization to create competitive multiplayer poker agents. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*, pp. 159–166. International Foundation for Autonomous Agents and Multiagent Systems, 2010.
9) グレーヴァ香子. 非協力ゲーム理論 (数理経済学叢書). 知泉書館, 5 2011.
10) B.W. Ballard. The*-minimax search procedure for trees containing chance nodes*. *Artificial Intelligence*, Vol.21, No.3, pp. 327–350, 1983.

⋆1 ナッシュ均衡の近似戦略を求めるアルゴリズムについては事前に戦略を計算し保存しておくものが多い．3 人 Texas Hold'em については CFR を 1ヶ月以上実行しており[8]，今回評価で用いた Leduc Hold'em では約 1 日半程度 CFR を実行した．