修　士　論　文

相手の抽象化による
多人数ポーカーの戦略の決定

Computing Strategies for Multiplayer Poker Games by Abstraction of Opponents

指導教員　近山 隆 教授

東京大学大学院工学系研究科
電気系工学専攻

氏 名　**37-116479　古居 敬大**

提 出 日　平成 25 年 2 月 6 日

概 要

本研究では不完全情報多人数ゲームのひとつであるポーカーでのコンピュータプレイヤの行動決定手法を提案する．

近年，不完全情報ゲームのポーカーにおいて，ゲーム理論に基づいたナッシュ均衡戦略の機会損失の最小化による近似計算手法が発展してきた．その成果もあり 2 人ゲームにおいては人間のチャンピオンクラスのプレイヤに勝ち越している．そのため現在では 3 人以上の多人数ゲームに関心が移りつつある．

3 人以上のプレイヤが存在する多人数ゲームでは 2 人ゲームでは生じなかった問題・課題が存在する．まずプレイヤの人数が増えるほど状態数が増大していくため，有効なナッシュ均衡戦略の近似計算が困難になってくることがあげられる．また多人数ゲームではナッシュ均衡の仮定である,「すべてのプレイヤが完全に合理的に行動する」ということが現実には合わない場合が多く存在することも課題としてあげられる．

提案手法では相手プレイヤを削減・抽象化した少人数のゲームを考慮し行動を決定する．これにより計算量を削減し高速なナッシュ均衡の近似的な戦略の事前計算が可能になると考えられる．対戦実験の結果，戦略の計算が困難になってくるプレイヤ人数が多いゲームでは特に，実際の人数よりも少ないが本来の人数に近いゲームの戦略が有効であることがわかった．

また相手プレイヤを削減することにより特定のプレイヤのみに着目した探索も可能になると考えられる．単純な行動をとるプレイヤとナッシュ均衡の近似戦略のプレイヤが混在する対戦実験を行った結果，単純なプレイヤを見つけ出して搾取的な探索をすることにより，単純なプレイヤから大きく報酬を得ることができ，ナッシュ均衡戦略よりも大きな報酬を得ている場合があることを確認した．

目 次

図 目 次

表 目 次

第1章　はじめに

1.1　背景

ゲームはルールが明確に定義されており，勝敗や報酬といった指標が与えられるため評価が行い易い，また古くから人々に親しまれているものであることからも，コンピュータゲームプレイヤは人工知能研究の一種のテストベッドとなってきている．

ゲームの中でも特に不確定不完全情報多人数ゲームという区分に属するゲームは，古くから研究が行われてきたチェスやオセロといった二人確定完全情報ゲームとは異なり，多くのプレイヤが存在し，それぞれが異なる情報を利用して個人の利得を最大化するという点で，現実世界の多くの諸問題をモデル化したものであると言える．

ポーカーは世界中で親しまれているトランプゲームの一つである．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．ポーカーは手札の優劣で勝敗を決定するゲームであるが，プレイヤの行動によって相手全員をゲームから降ろすことが出来れば勝ちになるなど，運の要素だけでなく戦略的な要素を多分に含んだゲームであるといえる．

ポーカーの分野におけるコンピュータゲームプレイヤの研究は，コンピュータチェスプレイヤの Deep Blue が人間のチェスの世界チャンピオンに勝利した頃である，1990 年代から徐々に行われてきており [1]，近年では AAAI や NIPS といった国際会議でもポーカーを題材とした研究が多く発表されている [2, 3, 4]．

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) について確率的に良い報酬が得られ，かつ相手に有利となる自らの情報をあまり与えないような，ナッシュ均衡 [5] 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．しかし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，手札の集約を行うといった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を求め，それを実際のゲーム中には利用することとなる．従来は線形計画問題を解くことで ϵ-ナッシュ均衡戦略を得ていた [6]．しかし近年ではオンライン学習の分野で注目されている機会損失 (Regret) の最小化手法を用いたナッシュ均衡の効率的な近似手法である，CounterFactual Regret minimization (CFR) [7] が提案された．その成果もあり，限定的なルールではあるが 2 人ゲームにおいてコンピュータプレイヤが人間のチャンピオンクラスのプレイヤに勝ち越すまでになった [8]．

ポーカーは不完全情報ゲームであるだけでなく，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．しかしコンピュータポーカープレイヤの研究対象は本来多人数ゲームであるにもかかわらず 2 人ゲームが中心であった．これは二人零和ゲームではゲーム理論に基づいたナッシュ均衡解が最適な戦略になるなど理論的な解析が行い易く，また計算量的な観点からも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しナッシュ均衡戦略の計算が難しくなるためであると考えられる．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 [9] であったり，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 [10] 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．また多人数ゲームでは人数が増えることによって状態数・計算量が大幅に増えるだけでなく，共謀や搾取といった 2 人ゲームでは存在しなかったプレイヤ同士の関係性も考慮に入れる必要があることも多人数ゲームの研究があまり行われていないことの原因であると考えられる．ナッシュ均衡戦略では相手プレイヤは全員が独立して完全に合理的に思考することを仮定しているため，これら共謀や搾取といったプレイヤが独立して行動していない状態について考慮できないだけでなく，単純な複数のプレイヤに対しても大きく負け越すといった問題が生じてしまう．

1.2 本研究の目的

本研究の目的は多人数ゲームで生じる，計算量の増大や仮定が現実的ではないといった，ナッシュ均衡戦略の問題の解決である．

参加人数に対して融通の効く多人数ゲームは，ゲームの参加人数が少し変化したところでゲームの本質は変化しないことが予想される．そのため多くのプレイヤが参加する状態空間の大きなゲームでも，それより少ない人数のゲームの戦略も有効に作用するのではないかと推察される．そこで本研究では多人数ゲームの「人数が多少変動してもゲームの性質は変わらない」という仮説の検証するとともにそれを利用したプレイヤの作成を行う．

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，ナッシュ均衡戦略を直接的に計算することが不可能な人数の多人数ゲームでも，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．また一方でナッシュ均衡戦略から外れた非合理な行動をとるプレイヤが存在する場合はそのプレイヤとの少人数のゲームを想定することでそのプレイヤにつけこんだ，搾取的な戦略をとることが可能になると考えられる．

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 [11] を 6 人まで参加できるように拡張したゲームを用いる．

1.3 本研究の貢献

本研究での貢献として，以下のようなものがあげられる．

- トイゲームではあるが，従来扱われてこなかった 4 人以上のプレイヤが参加するゲームでのナッシュ均衡戦略の近似計算を取り扱った．従来の研究では 2 人ゲームの拡張として取り扱われることが多く，分析もほとんど 3 人ゲームまでしか行われて来なかった．本研究ではトイゲームではあるが 6 人ゲームまでのナッシュ均衡の近似計算を行い，評価実験に用いている．
- 相手プレイヤの抽象化により，多人数ゲームでの計算量と戦略の非合理性の両面の解決を狙った．本研究で提案する少人数のゲーム戦略の適用は，人数が増加するに従い増える計算量の抑制と非合理的なプレイヤに着目した搾取を可能にしている．
- 評価実験により以下のことを確認した．
  - 少人数ゲームでのナッシュ均衡戦略がそれより多い人数でのゲームでも有効であること
  - ナッシュ均衡戦略の仮定が崩れる状況で相手プレイヤへの搾取戦略が有効であること

  このことから多人数ゲームで本来の人数とは異なる人数のゲーム戦略が活用出来るだけでなく，それらの戦略を報酬を最大化に有効に作用させられることも確認した．

1.4 本論文の構成

まず序論にて本研究の背景，目的について述べる．次に関連研究として今回扱うゲームであるポーカーゲームや，多人数ゲームや不完全情報ゲームでとられてきた行動決定手法について述べる．その後提案手法について説明した後，性能評価のための実験の内容とその結果を示し考察，最後にまとめを行う．

第2章　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．

ポーカーにはカードの配布（ディール）という偶然性が存在するが，それはある程度のゲーム回数を重ねれば平準化される.「2000 ゲームを越えれば実力の差がはっきりする [12]」とも言われている．

ポーカーの主な特徴として，以下があげられる．

- **ゲームそのものに勝つことが目的ではないこと**
  ポーカーは基本的にゲームを繰り返し行うため，1 ゲームで勝つことより最終的なチップ枚数を最大化することの方が重要である．

- **手の役自体に点数が無いこと**
  手牌を揃えて役を作る麻雀のようなゲームとは異なり，役自体には点数がない．そのため，たとえ良い役で勝ったとしても，相手が早々にゲームから降りてしまえば報酬は小さくなる．役そのものより対人競技性のほうが強いゲームであるといえる．

これらの特徴から降りる行動であるフォールドが戦略上とても重要な行動となり，フォールドと判断することが悪手になることはほとんどない．

ポーカーのルールには多くのバリエーションが存在する．大別するとドローポーカー，ホールデムポーカー，スタッドポーカーの三種類がある．本研究ではホールデムポーカーの代表的なゲームであるテキサスホールデムのトイゲームであるルダックホールデムを扱う．日本で広く行われているポーカーはカードの交換を行うドローポーカーであり，本研究の題材として用いているホールデムとは若干性質の異なるものである．

2.1　テキサスホールデム

テキサスホールデム (Texas Hold'em) [2]はポーカーの一種である．テキサスホールデムの主な特徴としては，

[1] ワンペア，ツーペアなど
[2] 単にホールデムとも

表 2.1: ポーカーの役

名称	説明
ロイヤルフラッシュ	同じスートで T〜A まで連続したカード 5 枚
ストレートフラッシュ	同じスートで連続したカード 5 枚
フォー・オブ・ア・カインド	数字が同じカード 4 枚
フルハウス	同じ数字 3 枚と，同じ数字 2 枚の組み合わせ
フラッシュ	同じスートのカード 5 枚
ストレート	スートに関係なく数字が連続した 5 枚
スリー・オブ・ア・カインド	数字が同じカード 3 枚
ツーペア	「同じ数字 2 枚」が 2 組
ワンペア	同じ数字 2 枚
ハイカード	上記以外のカード 5 枚

- **コミュニティカード** 各プレイヤが共有して使用するカードが存在する．
- **ベットラウンド制** ベットにラウンドを設け，段階的にベットを行なっていく．

などがある．手の役としては一般的なポーカーゲームとして知られている表 2.1 の役がある．

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するかの決断を行う．

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．例として図 2.1 をあげる．プレイヤ 1 は ♠ 7 から ♠ A までの下の 6 枚，プレイヤ 2 は ♡ 2 から ♠ 6 までの上の 6 枚が現在の手札となる．カードを交換するドローポーカーとは異なり，使用出来るカードはラウンドを経るにつて増えるため，ゲームに参加し続ければ手札の役は強くなっていく一方で，決して弱くなることはない．しかし相手プレイヤも同様に強くなるので，自らの手札の将来性について相手の行動を見ながら判断していく必要がある．

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．

2.1.1 ゲームの流れ

具体的なゲームは以下のような流れで行われる．

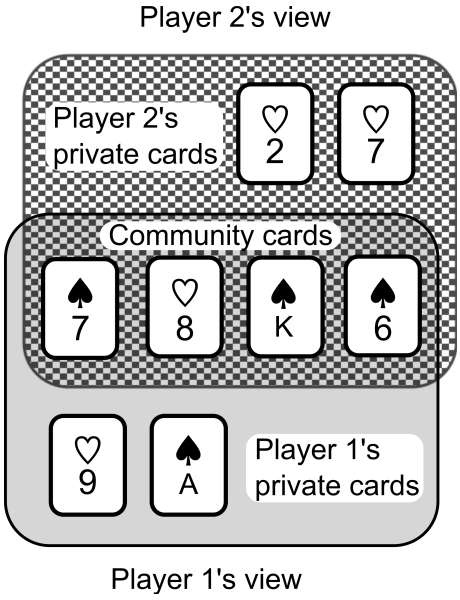


図 2.1: コミュニティカード

1. 参加費となるアンティやブラインドベットを場に出す
2. ホールカード (手札) が配られる
3. ベットラウンド
   - **プリフロップ** 手札と相手のベットアクションをもとに各自ベットアクションを決定する
   - **フロップ**，**ターン**，**リバー** 各ラウンド開始時に 1 枚ずつコミュニティカードを公開し，各自ベットアクションを決定する

各ベットラウンド中の具体的なベットアクションの選択肢は以下の 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 2.2 に示した．

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 2.3 のようにプリフロップからリバーまでの 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．スートに関しての優劣はないため引き分け・複数人の勝者も発生するがその場合報酬は勝者の間で等分される．

表 2.2: 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}	$\sim 10^n$

2.2 ルダックホールデム

ルダックホールデム (Leduc Hold'em) [11] 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 2.2 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（Ante, アンティ）を賭け，報酬として -13 から $+13(n-1)$ 枚[3]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．

2 人ゲームの場合，カードの不確定性を除いた行動に関するゲーム木は図 2.4 のようになる．プレイヤの手札を観測できない第三者からゲームを見た場合のゲーム木（Public Tree[13]）の非終端ノードの数は 2 人ゲームでは 85，3 人ゲームでは 1120 である．以降人数が増えるに従いおよそ 10 倍になっていく．

[3] n はゲームへのプレイヤの参加人数， [11] では $n=2$，後に $n=3$ で実験が行われている．表 2.2 の山札の枚数については $n=2,3$ から一般化を行なった．

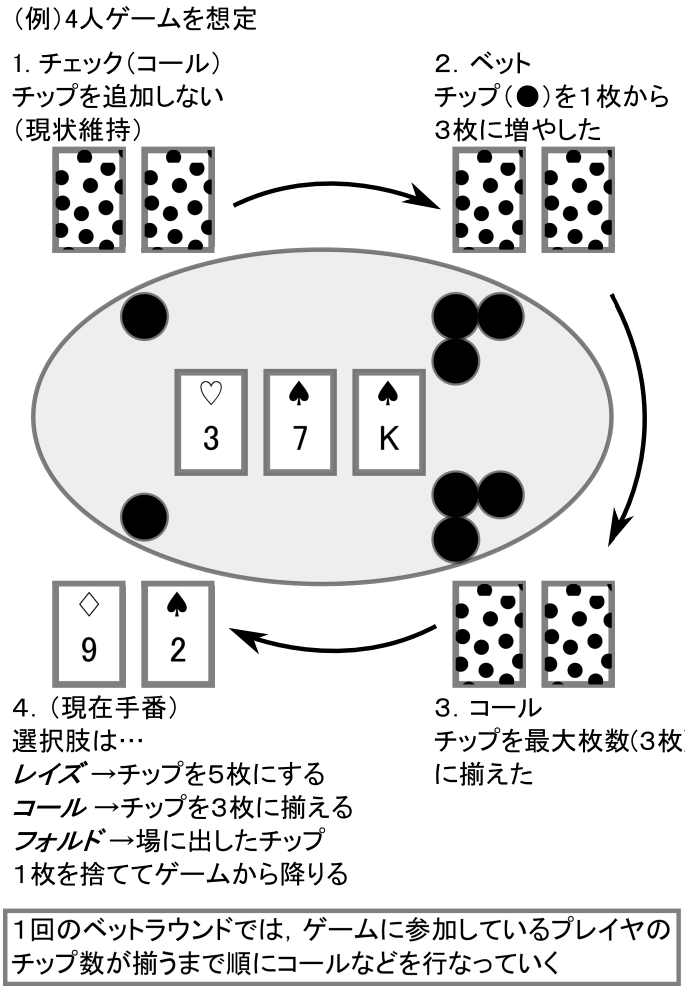


図 2.2: ベットラウンドでのプレイヤの行動

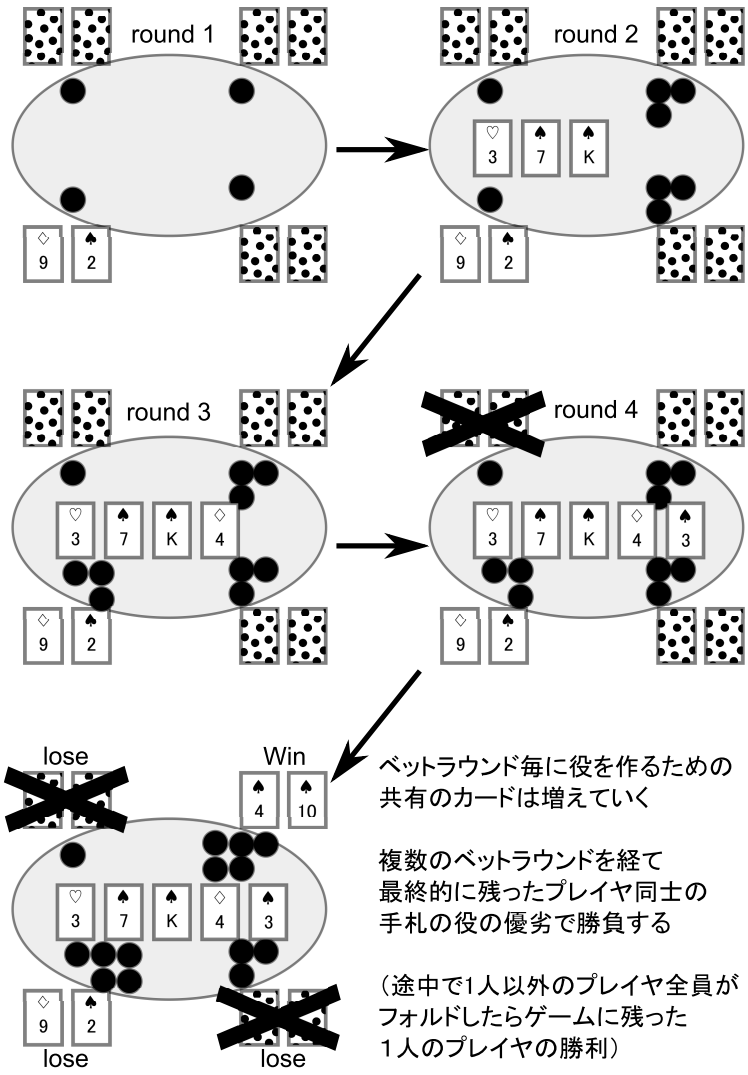


図 2.3: テキサスホールデムのゲームの一連の流れ

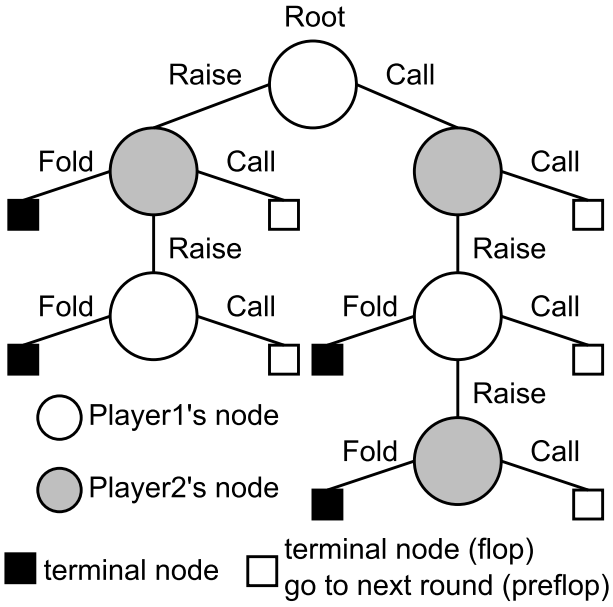


図 2.4: 2-player Leduc Hold'em

第3章　関連研究

3.1　不完全情報ゲームと完全情報ゲーム

完全情報ゲームとは全てのプレイヤが同じ情報を共有するゲームである．代表的な完全情報ゲームとしてオセロやチェス，囲碁や将棋などがある．

完全情報ゲームの行動決定では現在の局面の良し悪しを判断する評価関数とミニマックス探索とを組み合わせたものが一般的である．また近年ではモンテカルロシミュレーションに探索の概念を導入したモンテカルロ木探索が提案され，その汎用性の高さから多くのゲームの行動決定で利用されている．

一方で不完全情報ゲームとは各プレイヤが異なる情報を基に行動を決定するゲームである．プレイヤに手札が存在するトランプのゲームなどが例としてあげられる．

不完全情報ゲームはオセロやチェスといった確定完全情報ゲームとは異なり，現在局面の情報全てが特定のプレイヤは取得できず，図 3.1 のように現在状態は情報集合のいずれかのノードにあるのだが，各プレイヤはどのノードにいるのかがわからない．よって単一のプレイヤでは現在局面を一つに絞ることが不可能であるため，完全情報ゲームでたびたび用いられている評価関数とゲーム木探索による行動の決定手法やモンテカルロ木探索といった手法を直接的には利用できず，利用するのであれば，相手の戦略に何らかの仮定をするなどして，現在局面の推定を行わなければならない．

3.1.1　戦略

本論文では戦略 (strategy) はゲームの各局面でどの行動を取るかを決めるための確率の組み合わせの集合（確率表）のことを指す．不完全情報ゲームでは現在局面を一意に決定できないため，情報集合毎に行動とその確率が与えられる．

戦略には大きく分けて，一連のゲームを通じて戦略が変化しない静的戦略と，ゲームの状況や相手プレイヤの行動に応じて戦略を適応させていく動的戦略とがある．

戦略の分析においては最適応答 (Best Response) の計算がたびたび用いられる [10, 13, 2]．最適応答は相手戦略を既知とした場合の最善の戦略であり，これにより戦略の搾取されやすさなどを測定することが可能である．

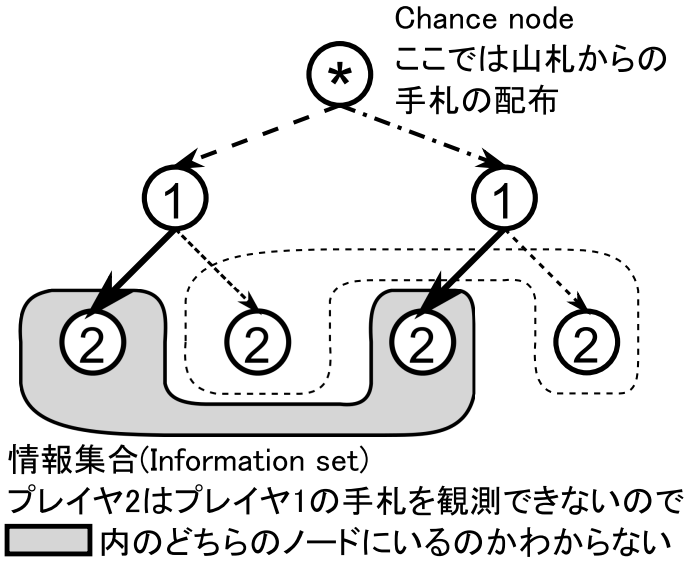


図 3.1: 不完全情報ゲームと情報集合

3.2 コンピュータポーカープレイヤの研究

3.2.1 ルールベース，数式ベースのプレイヤ

ルール記述やある種の公式を基に行動を決定する知識ベースのプレイヤは古くから研究されており，商用のコンピュータプレイヤとしても実装されている [14, 15, 16].

しかしながらルールベースによるコンピュータプレイヤの強さは，製作者の熟達さに大きく依存する．またルールを増やしたり複雑なルールを導入すればコンピュータプレイヤが強くなる可能性はあるが，ルール同士の衝突を防ぐことが困難になってくるといった問題もある．

3.2.2 モンテカルロ法

モンテカルロ法は乱数を用いてなんらかの近似値を推定する手法である．

ゲームにおいては不確定な要素やプレイヤの行動をランダムに決定して得られる結果・報酬の平均値が最良となる行動をとるものが考えられる．近年では木探索を導入したモンテカルロ木探索 [17] が提案され，多くのゲームで利用されている．

現在状態が確定できない不完全情報ゲームのポーカーでも相手戦略のモデル化などを利用することで現在状態を推定することでシミュレーションを可能にするといった研究が行われている [18].

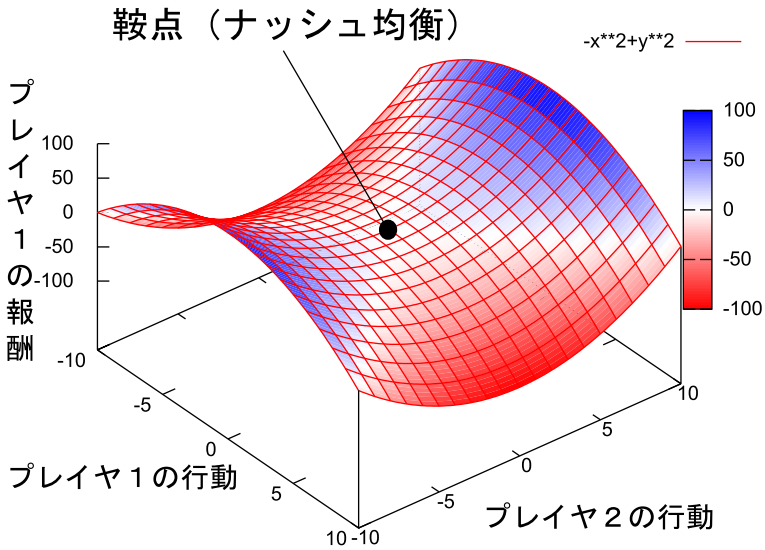


図 3.2: 鞍点

表 3.1: 囚人のジレンマ

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

3.2.3 ナッシュ均衡の近似計算

3.2.3.1 ナッシュ均衡

ナッシュ均衡 [5] 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，という均衡した状態を指す．2 人零和ゲームで生じるナッシュ均衡は図 3.2 の鞍点のような状態である．

有名なナッシュ均衡の例としては表 3.1 の囚人のジレンマがある．

囚人のジレンマは 2 人非零和ゲームで生じる，互いに利己性を追求することで報酬が減少するという現象である．一見お互いにとって望ましいのはお互いに協力して黙秘を貫く組み合わせであるが，ナッシュ均衡の考えでは自己の利益（刑期の軽減）しか考えていないため，相手が協力・裏切りどちらでも報酬が増える（刑期が減る）裏切りを選択してしまい，お互いにとって不幸せな状態になる，というジレンマである．

ナッシュ均衡は有名な「囚人のジレンマ」の例のように，必ずしも，戦略を変更することでこれ以

上誰も損をすること無く利得を上げられるパレート最適な状態では無い．しかし非協力零和ゲームであるポーカーなどのゲームでは，各プレイヤが協調や共謀をしない限りにおいては，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 [6].

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[1].

3.2.3.2 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 [8].

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．

$$R_i^T(I,a) = \frac{1}{T}\sum_{t=1}^{T} \pi_{-i}^{\sigma^t}(I)(u_i(\sigma^t|_{I\to a}, I) - u_i(\sigma^t, I)) \tag{3.1}$$

図 4.3 のように，ここで i は I での手番プレイヤ，$\pi_{-i}^{\sigma^t}(I)$ 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[2]，$u_{i,I}$ はノード I でのプレイヤ i の平均報酬であり，$u_i(\sigma^t|_{I\to a}, I)$ 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．$u_i(\sigma^t|_{I\to a}, I) - u_i(\sigma^t, I)$ 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 3.2 によって抽出してそれを R^+ とし，式 3.3 のように次回イテレーションで利用する戦略 $\sigma_i^{T+1}(I)(a)$ へ用いる．

$$R_i^{T,+}(I,a) = \max(R_i^T(I,a), 0) \tag{3.2}$$

[1] One-card poker http://www.cs.cmu.edu/ggordon/poker/
[2] 手番が i のノードでの遷移確率は 1 とする

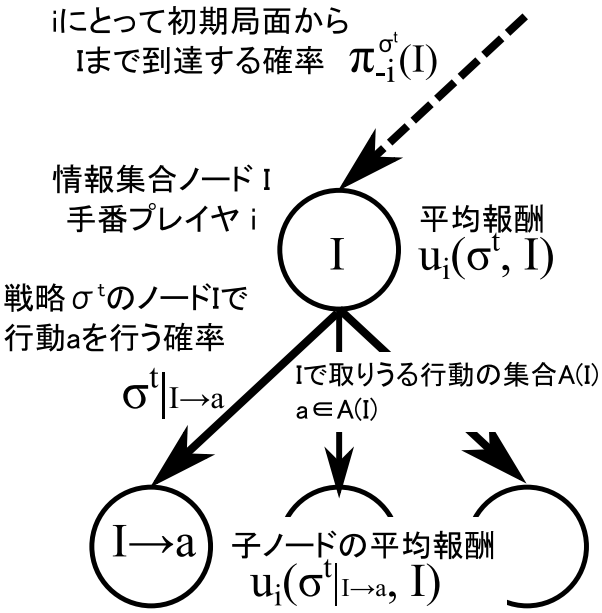


図 3.3: 情報集合ノードとそこでの戦略と平均報酬

$$\sigma_i^{T+1}(I)(a) = \begin{cases} \frac{R_i^{T,+}(I,a)}{\sum_{a\in A(I)} R_i^{T,+}(I,a)} & \text{if} \sum_{a\in A(I)} \\ \frac{1}{|A(I)|} & \text{otherwise.} \end{cases} \tag{3.3}$$

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 3.4 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 [19] 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．

$$\overline{\sigma}_i^T(I)(a) = \frac{\sum_{t=1}^T \pi_i^{\sigma^t} \sigma^t(I)(a)}{\sum_{t=1}^T \pi_i^{\sigma^t}(I)} \tag{3.4}$$

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 [10]．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．

3.2.3.3 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．

3.2.3.4 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が判断できればよいため，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 [20].

3.2.3.5 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 [21] といった選択肢の抽象化・集約手法もある.

3.2.3.6 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 [22] 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.

3.2.4 相手のモデル化

現在状態を推定するための相手の戦略を予想することは有効である．しかしながら複数の戦略を切り替えたりすることでモデル化を防いだりすることも考えられるし，相手のモデル化には相手モデルの精度・適応性だけでなく，そのモデルへの学習速度や相手の戦略が切り替わった時に早く適応できることが必要となる.

古くは，度数表（Action Frequencies）を利用したもの [14] や，ニューラルネットワークを用いることでモデルの表現力を増加させたもの [23] があり，ニューラルネットワークを利用したものは実験で 70 %から 80 %の精度が出たとの報告がある.

近年ではナッシュ均衡戦略からずらすアプローチで相手の戦略に適応しつつ，ナッシュ均衡戦略的な行動を取らせるなどの手法も提案されている [24].

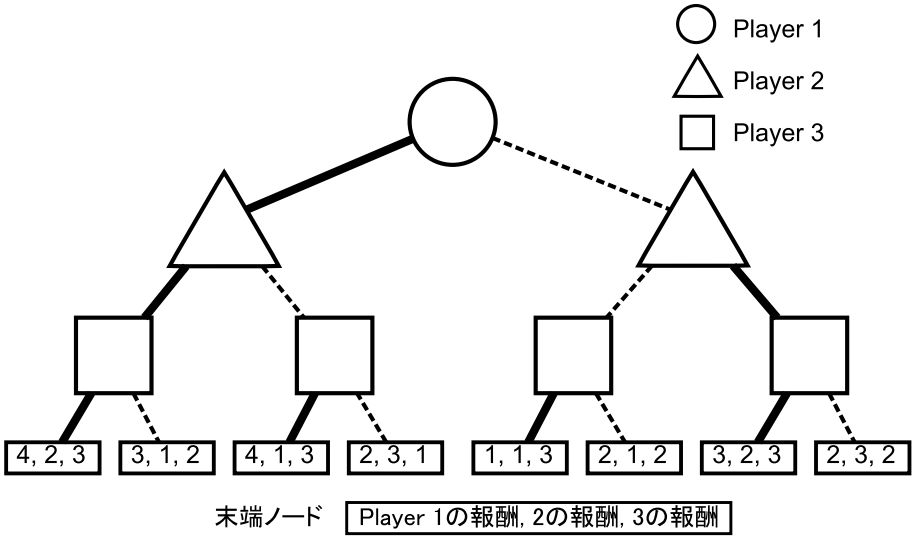


図 3.4: $\mathrm{Max}^{\mathrm{n}}$

3.3 完全情報多人数ゲームのコンピュータゲームプレイヤの研究

不完全情報多人数ゲームにおけるゲームプレイヤを扱った研究は，モンテカルロ木探索や CFR といった 2 人ゲームでの手法の適用といったものにとどまっている．

一方で完全情報多人数ゲームのコンピュータプレイヤの研究は，２人完全情報ゲームにおけるミニマックス探索の拡張として行われてきている．

3.3.1 $\mathrm{Max}^{\mathrm{n}}$

$\mathrm{Max}^{\mathrm{n}}$ 探索 [25] は各プレイヤが自身の報酬・効用を最大化するという仮定のもとに行われる，完全情報多人数ゲームにおける基本的な探索である．図 3.4 のように各プレイヤのノードでは子ノードのそのプレイヤの報酬・評価値が最大となるノードを選択するように行動を行い，親ノードへ評価値を伝搬させていく．

3.3.2 Paranoid アルゴリズム

Paranoid アルゴリズム [26, 27] は，相手全員が自プレイヤの報酬を最小化するという仮定をおいた探索手法である．

図 3.5 ではプレイヤ 1 については自分の報酬・評価値を最大化し，他のプレイヤはプレイヤ 1 の報酬・評価値を最小化しようとしている．基準となる評価値が自プレイヤのもののみになることから，

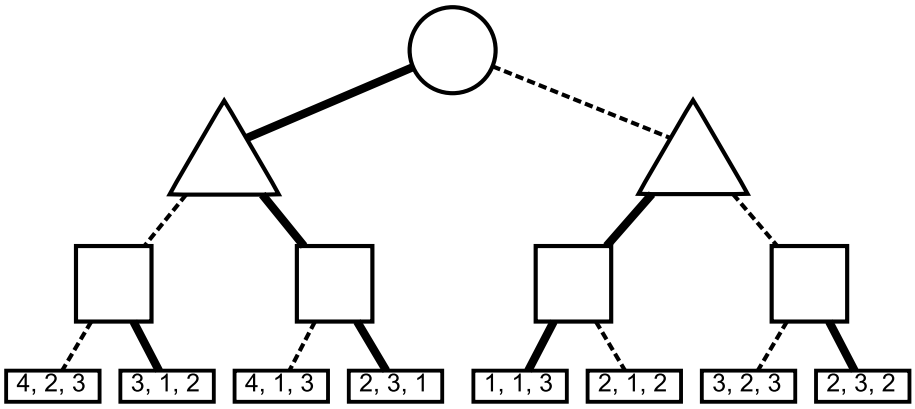


図 3.5: Paranoid アルゴリズム

探索はミニマックス探索と同様のものとなり，Alpha-Beta 枝刈りと同様の探索木の枝刈りが可能となり，より高速な探索が可能となる．

しかしながら，相手全員が自プレイヤの報酬を最小化するという仮定は現実的には成立しない場合が多く，結果として消極的な守りの行動をとってしまう可能性がある．

3.3.3 Best-Reply Search

Best-Reply Search [28] は多人数ゲームにおける探索手法の一つである．Best-Reply Search も Paranoid アルゴリズム同様に相手プレイヤの戦略は自プレイヤの報酬を最小化するものという仮定を行う．しかし探索では相手プレイヤの行動を順に探索するのではなく，相手プレイヤの行動の一つを相手プレイヤの行動とみなした探索を行う．

自プレイヤは報酬を最大化し，相手プレイヤは自プレイヤの報酬を最小化する意味では Paranoid アルゴリズムと同様であるが，図 3.6 のように図 3.4 や 3.5 とは異なり，３人のプレイヤが順番に行動することを想定せず，相手プレイヤのうちの１人が自分の報酬を最小化するある種の２人ゲームを想定している．

このような探索は時にゲームルールに対して矛盾を生じさせるものになるかもしれないが，探索の速さや，相手プレイヤに対する仮定の緩和などにより従来の探索手法に勝ち越すことが可能となっている．

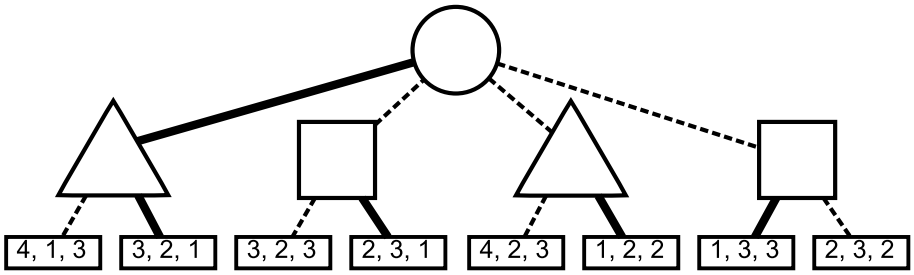


図 3.6: Best-Reply Search

第4章　提案手法

多人数ゲームでのナッシュ均衡戦略には以下のような問題があると考えられる．

1. **計算可能性** そもそも状態数が多すぎて計算出来ない．
2. **非合理なプレイヤによるナッシュ均衡の仮定の崩壊**仮にナッシュ均衡戦略がとれたとしても，全員が同様にナッシュ均衡戦略を取ることを仮定した最適戦略であるため，仮に複数のプレイヤが想定していない非合理な行動を取ると報酬が減少してしまう可能性がある．

参加人数に対して「融通がきく」ような多人数ゲームでは，図 4.1 のようにゲームへの参加人数が少し増減したと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．また一方で特定のプレイヤに注目して行動を決定するような場合においても，注目していないプレイヤの行動の有無は重要でない場合が多いことが予想される．

そこで不完全情報多人数ゲームにおいて，実際のゲームよりも少ない参加人数のゲームを想定して行動を決定する手法を提案する．少人数のゲームを想定することのねらいとしては以下の 2 つがあげられる．

1. 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も，それよりも少ない人数での戦略を適用することでナッシュ均衡的な戦略を行うことが可能になる．
2. ナッシュ均衡的な戦略から大きく外れたプレイヤが存在する時にそのプレイヤに着目した探索や行動決定が可能になる．

本研究では，共通した「少人数へのゲーム化」を軸として，2 つ異なる戦略を提案する．概要をまとめて図 4.2 に示した．まず 4.1 節で共通事項である，少人数のゲーム化について説明する．その後 2 つの戦略（戦略 A，戦略 B）をそれぞれ順に説明していく．

4.1　相手プレイヤの行動の除外による少人数のゲームへの変換

特定プレイヤ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そのまま適用できるようになる．

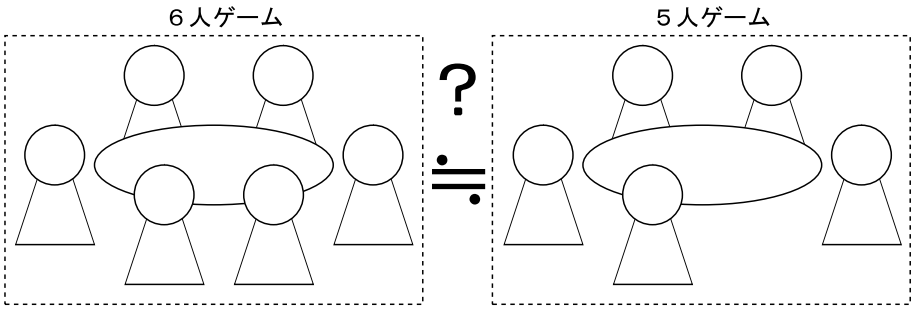


図 4.1: ゲームの本質は変化しない？

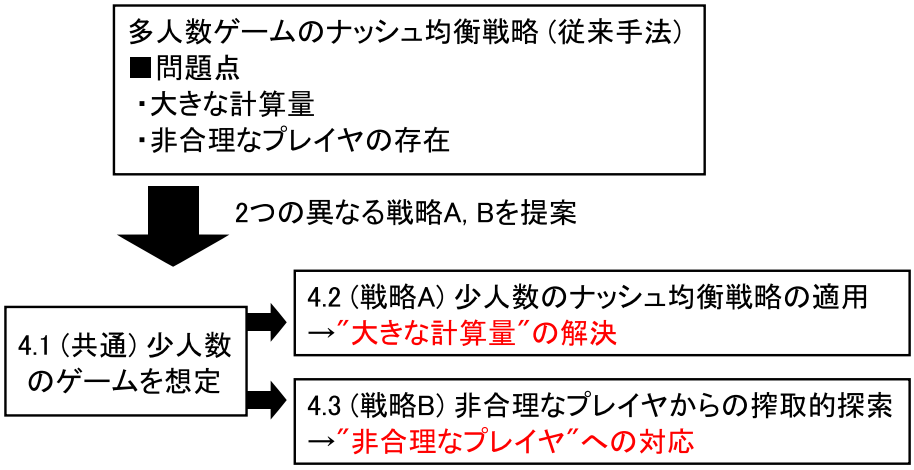


図 4.2: 提案手法の概要

プレイヤ行動の除外については今回は図 4.3 のような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な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 人づつ除外していく
2. フォルドしたプレイヤ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．

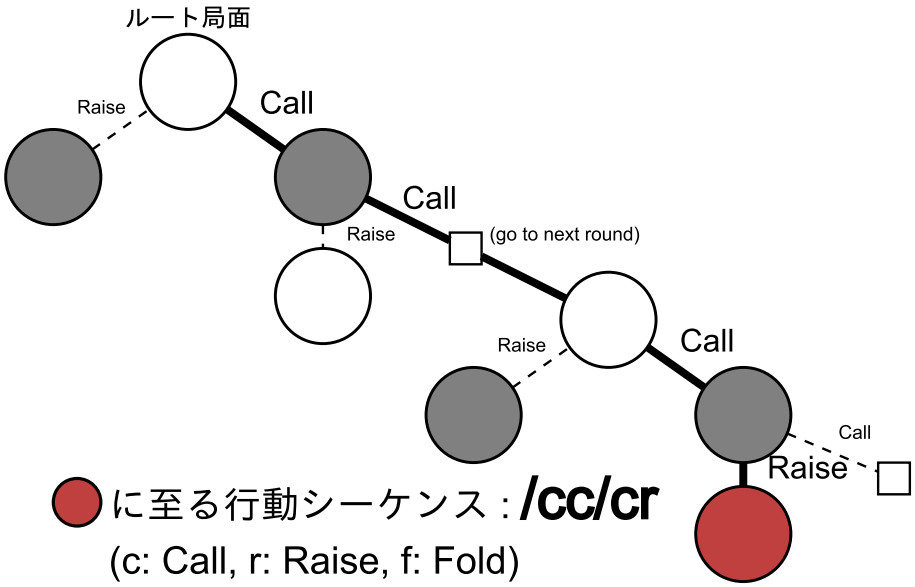


図 4.3: 行動シーケンス

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 Algorithm4.1 の通りである．削減する特定のプレイヤ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，Algorithm4.1 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 4.5 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．

4.2 (戦略 A) 少人数のゲームでの ϵ-ナッシュ均衡戦略の適用

図 4.6 のように前述の行動シーケンスの変換によって n 人ゲームから m 人ゲーム $(n > m)$ への変換を行い，m 人ゲームにおける ϵ-ナッシュ均衡戦略を n 人ゲームにて適用する．こうすることで実際のゲームで ϵ-ナッシュ均衡戦略を求めるよりも少ない計算量で，より精度の良い ϵ-ナッシュ均衡戦略を利用できることが期待される．

ここでは「全てのプレイヤが合理的な行動を取っている」というナッシュ均衡の仮定から，どのプレイヤを除外しても残るプレイヤ全てが合理的なプレイヤであると想定する．そのため行動を除外する特定のプレイヤについては自プレイヤを除いたプレイヤのうち，等確率でランダムに選択するものとする．

また適用する少人数ゲームでの ϵ-ナッシュ均衡戦略については事前に CFR によって計算したものを用いる．

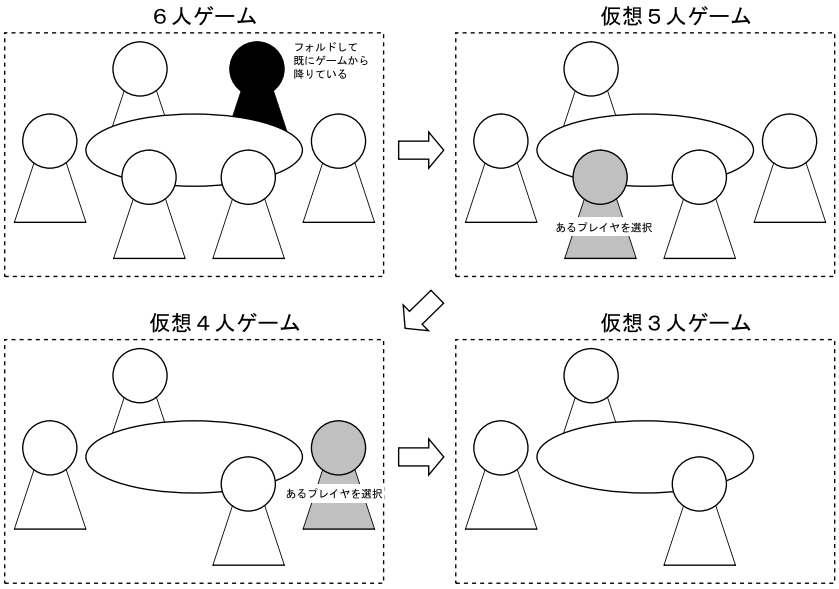


図 4.4: 変換の方針

4.3 (戦略 B) 非合理なプレイヤからの搾取的探索

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をあまり考慮しなくても，最終的な報酬は高くなる事が考えられる．特にナッシュ均衡戦略から大きく外れたプレイヤが存在するようなときにはそのプレイヤに着目した行動を取ることが有効であると考えられる．

そこで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し，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考え，そのプレイヤに対して搾取的な行動を取ることを提案する．

1 ゲーム中における提案手法全体の流れとしては以下のようなものとなる．

1. **相手のモデル化**
   相手の過去の行動の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

2. **搾取相手の選択**
   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

3. **選択した相手のみに着目した搾取的な探索**
   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

図 4.5: 提案手法の適用例

4.3.1 相手のモデル化

4.3.1.1 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，今回は特にナッシュ均衡戦略から大きく外れているであろう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．

4.3.1.2 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

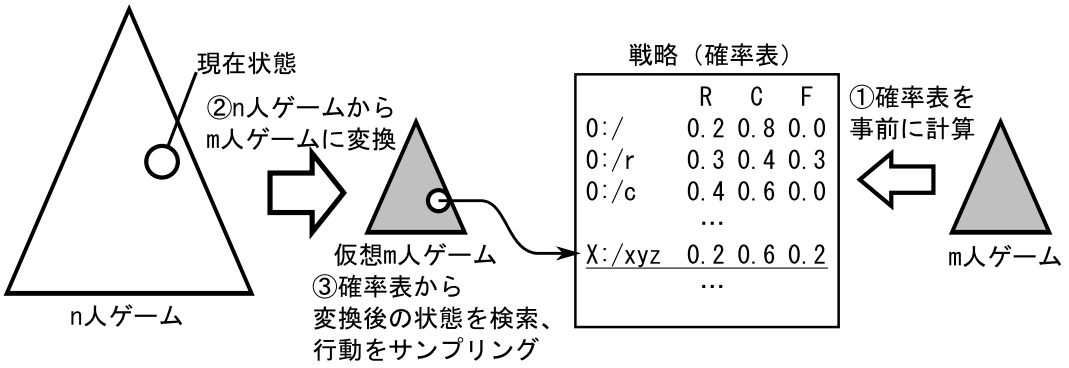


図 4.6: 提案手法の適用例

- ラウンド（プリフロップ，フロップ）
- 場のカード (コミュニティカード) 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）

によって分類し，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，

$$p_{opp}(n,a) = \frac{c_{n,a}}{\sum_{a'} c_{n,a'}} \tag{4.1}$$

とした．ただし c を頻度，n を状況，a を行動とする．このモデル化は，

1. 観測していない状況のモデル化ができない
2. 現れにくい状態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，単純に等確率で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は相手のプライベートカードを観測できないため，特定のプライベートカードにおけるフォールド回数を，

$$c_{n,a=f} = c_{n',a=f}/N_{card} \tag{4.2}$$

というように単純に Leduc Hold'em カードの種類数 N_{card} で割ることで近似している．

4.3.2 搾取対象の選択

予め算出しておいた ϵ-ナッシュ均衡戦略の確率表[1]を用いて，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，以下のようなナッシュ均衡からの仮想的な距離を表す L を定義する．

$$L = \frac{\sum_{n=1}^{N} -\log(P_{\epsilon Nash}(n,a))}{N} \tag{4.3}$$

これは ϵ-ナッシュ均衡戦略が現在までの行動を取る確率の対数に–1 を掛けたものの平均である．この L は ϵ-ナッシュ均衡戦略を取るプレイヤの場合については 0 に近い値になり，逆に ϵ-ナッシュ均衡戦略から離れる戦略の場合には大きな値を取る．今回は L 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．

4.3.3 特定の相手のみに着目した搾取的戦略の実行

特定の相手（搾取対象）のみの行動に着目するために，それ以外の相手プレイヤの行動は原則除外して，仮想的な 2 人ゲーム木上を探索する．除外は 4.1 節のように行動シーケンスの変換によって，搾取対象と自プレイヤ以外の行動を除外するものとする．

そのシーケンスを用いて仮想的な 2 人ゲーム木上で Fig.4.7 のような Expectimax 探索 [29] を行う．この探索は自プレイヤの行動ノードで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．

[1] この確率表は厳密なナッシュ均衡戦略でなくて良く，短時間の CFR 実行によって生成する．後述の実験では ϵ-ナッシュ均衡戦略を取るプレイヤとの対戦を行うが，そこで用いる相手プレイヤの戦略を生成するための CFR の実行時間よりははるかに小さいものを用いる．

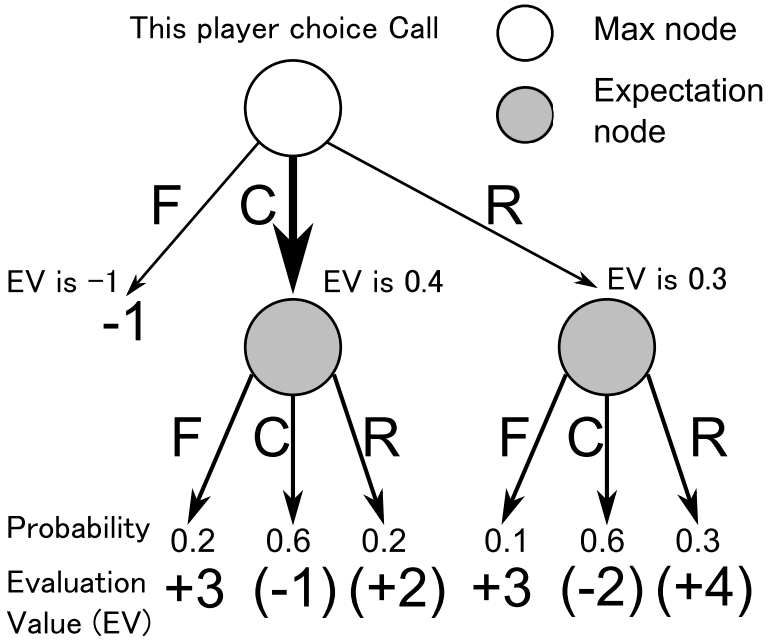


図 4.7: Expectimax 探索

Algorithm 4.1 シーケンスの変換を用いた 1 プレイヤ削減

```
//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 ⇐ 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=除外対象 then
    //除外対象プレイヤの行動はレイズ以外反映しない
    if s[i]=レイズ then
      next_raise ⇐ true
    end if
  else
    //その他のプレイヤの行動は基本的にそのまま
    if s[i]=コール ∧ next_raise=true then
      s_new[j]⇐ レイズ
      next_raise ⇐ false
    else if s[i] が s_new[j] では非合法手 then
      s_new[j]⇐ コール
    else
      s_new[j]⇐ s[i]
    end if
    j ⇐ 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 ⇐ 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] ⇐ レイズ
  j ⇐ j + 1
end while

return s_new[]
```

第5章　評価実験

本章では提案手法の評価を行う．実験は大きく分けて以下の 2 つを実施する．

1. **戦略 A の検証** 少人数ゲームでのナッシュ均衡戦略の有効性の検証
2. **戦略 B の検証** 特定の単純なプレイヤに限定した探索の有効性の検証

5.1　共通の実験設定

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．

5.2　(戦略 A) 少人数のゲームでの ϵ-ナッシュ均衡戦略の適用 ϵ-ナッシュ均衡戦略同士の対戦実験

4.2 節の少人数ゲームでのナッシュ均衡戦略の適用についての評価を行う．短時間で収束すると考えられる少人数ゲームでの戦略の有効性を検証するため，戦略を生成するための CFR の実行時間を一定時間に限定したプレイヤ同士での対戦を行う．

5.2.1　実験設定

ゲームは 3 人ゲームから 6 人ゲームまでそれぞれ行う．n 人のゲームでのプレイヤの組み合わせは，取りうるプレイヤの組み合わせが多いため本実験では，提案手法を取るプレイヤが 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を用いた．提案手法のプレイヤは 4.1 節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-ナッシュ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の選択については前述の方針の通り，フォルドしているプレイヤから選択していき，想定人数まで達しなかった場合は自分以外のプレイヤからランダムに選択するものとする．

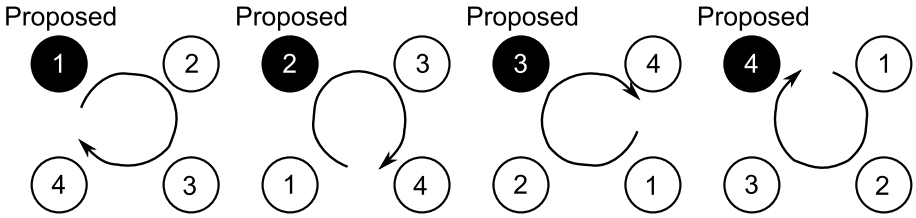


図 5.1: プレイ順序

表 5.1: 同一の時間 (24 時間) 実行した場合のプレイヤ人数ごとの CFR のイテレーション回数

人数	回数
2	656070042
3	44125140
4	3007959
5	334309
6	37682

対戦については，事前に各プレイヤの手札をランダムに 20 万ゲーム分生成し，プレイヤ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の平均報酬を用いる．

提案手法のプレイヤ以外のプレイヤ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 5.1 のように提案手法プレイヤの順序を変えた n 通りについて行う[1]．

提案手法，比較手法のプレイヤ共に用いる戦略 (確率表) は，共通して CFR によって生成された確率表を利用するものとする．[2]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などは行なっていない．CFR の計算時間は各人数毎に 24 時間[3]で固定とし，対戦実験では計算途中の戦略表を適宜出力し[4]，それらの組み合わせについて対戦を行った．

5.2.1.1 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人数の関係を表 5.1 に示す．プレイヤの人数が 1 人増えるたびにイテレーション回数がおよそ 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．

[1]5 人ゲームなら 20 万× 5 通りの 100 万回のゲームを行う．
[2]例えば 5 人ゲームの実験で相手プレイヤとして用いた戦略は 6 人ゲームの実験では提案手法プレイヤが利用するものとする．
[3]Intel Core i3 530 2.93GHz 1 thread にて
[4]例えば 6 人ゲームではイテレーション 1000 回，2000 回，3000 回…という具合に一定回数毎に確率表を出力した．

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特定のノードを重点的に更新していく手法 [30] などがあるが，多人数ゲームでは更新されるノードがスパースになりすぎると考えたため今回は行なっていない．

5.2.2 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[5]，高いほうが良い結果となる．平均標準偏差は 0.005 程度である．

5.2.2.1 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 5.2,5.3 に示す．右に行けば行くほど比較プレイヤの 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に対する提案手法の平均報酬となる．そのためグラフは全体として右下がりになっていることがわかる．

5.2.2.2 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 5.4,5.5 に示す．

例えば図 5.5(b) では 3 人 CFR や 4 人 CFR を利用した提案手法プレイヤは比較プレイヤの 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．

5.2.3 考察

5.2.3.1 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 5.2(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 5.2(b) から 5.3(b) までより 1 人減らした $n-1$ 人の CFR を利用したプレイヤを用いても平均報酬はほとんど± 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の戦略は図 5.4 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．

[5]単位は ante（参加費としてのチップ 1 枚），すなわちチップ枚数，参加費・報酬は 2.2 節を参照

表 5.2: 初期局面でレイズを取る確率

ランク	2-CFR	3-CFR	4-CFR	5-CFR	6-CFR
0	0.323	0.141	0.044	0.028	0.002
1	0.260	0.134	0.046	0.048	0.004
2	0.059	0.085	0.043	0.080	0.005
3	0.002	0.053	0.003	0.069	0.077
4	0.136	0.012	0.005	0.047	0.127
5	0.191	0.173	0.086	0.242	0.353
6	0.695	0.483	0.196	0.403	0.428

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 5.2(b), 5.3(a) 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 5.3(b) のプレイヤが想定より 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 5.5(b) 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 5.5(b) 右端や図 5.3(b) のように 4 人 CFR と遜色ない成績を収めていることからも，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．

また 6 人ゲームについて相手プレイヤの戦略の CFR 実行時間を 72 時間まで延長して行った対戦結果について図 5.6 に示す．十分に時間をかけることで本来の人数のゲームで生成された戦略をとるプレイヤの報酬が大きくなり，提案手法プレイヤの報酬が減少していくことがわかる．24 時間を超えて 3 人 CFR の報酬が負になっていることからも，時間をかければ 4 人,5 人ゲームの CFR を利用した提案手法プレイヤの報酬もいずれは負になると考えられるが，24 時間の 3 倍の 72 時間実行しても提案手法プレイヤの報酬は負になっていない．

6 人ゲームで 5 人ゲームの戦略を適用させた場合について，それぞれ戦略の計算時間を 24 時間までの間で変化させたときの提案手法の報酬について図 5.7 に示す．この時間内では同一の学習時間であれば常に提案手法が勝ち越していることがわかる．負け越しているのは図 5.5(b) のような非常に短い時間の学習時間の場合のみである．

5.2.3.2 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表 5.2 と図 5.8 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

表 5.3: 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．平均標準偏差は 0.01 以下である．

対戦組合せ	提案 1	比較プレイヤ 2	3	4	5	6
2 vs 3	-0.085	0.033	0.051			
2 vs 4	-0.218	0.054	0.080	0.083		
3 vs 4	-0.013	0.007	0.007	-0.001		
2 vs 5	-0.321	0.066	0.084	0.086	0.085	
3 vs 5	-0.022	0.012	0.011	0.001	-0.002	
4 vs 5	0.006	0.000	-0.005	0.000	-0.001	
2 vs 6	-0.342	0.048	0.065	0.074	0.066	0.089
3 vs 6	0.006	0.014	-0.003	-0.005	-0.003	-0.009
4 vs 6	0.070	-0.016	-0.009	-0.011	-0.015	-0.019
5 vs 6	0.069	-0.022	-0.008	-0.021	-0.009	-0.008

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．

この局面では 4 人ゲームの戦略は他の戦略にくらべ大きく消極的であり傾向がことなるものであるが，この 4 人ゲームの戦略も図 5.5(b) 実験で有効に作用していることがわかる．

5.2.3.3 比較プレイヤの報酬のプレイ順序による依存性

プレイヤの順序による報酬の違いについて表 5.3 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 (1) の直後のプレイヤ (2) についてはほとんどの場合で提案手法のプレイヤとは報酬の正負が反転しており，提案手法のプレイヤ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ほど提案プレイヤ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．

5.2.3.4 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 5.2, 5.4 のようなグラフの結果から人数が増えてもイテレー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．このナッシュ均衡への収束性の保証がない CFR の収束点が実際のゲームにおいて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．

5.3 (戦略 B) 非合理なプレイヤからの搾取的探索 特定のプレイヤからの搾取実験

5.3.1 実験設定

4.3 節の特定に対する搾取的な探索を行う提案手法プレイヤと，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とでルダック・ホールデムを繰り返し行う．

プレイヤの構成人数や順序については，提案手法プレイヤ 1 人，単純な行動をとるプレイヤ 1〜2 人，残りは ϵ-ナッシュ均衡戦略のプレイヤとする．それぞれのプレイヤは残りのプレイヤがどんな戦略を取るのか事前には知らされておらず，提案手法プレイヤとしてはどのプレイヤが単純な行動のプレイヤかを判断・選択する必要がある．

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. **提案** 提案手法
2. **ϵ-Nash** ϵ-ナッシュ均衡戦略（CFR により生成）
3. 単純な行動を取るプレイヤ
   - **RAISE** 常にレイズ（レイズができないときはコール）
   - **CALL** 常にコール
   - **RAND** ランダム

ランダムな行動を取るプレイヤ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．なお単純な行動をとるプレイヤのうち後方に (p) とつくプレイヤについては確率 $1-p$ で ϵ-ナッシュ均衡戦略を取るものとする．

対戦はプレイヤの組み合わせ 1 つにつき，プレイヤの順序ごとに 2000 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．

表 5.4: L の基準にした戦略の CFR イテレーション回数

人数	回数
3	39783
4	3533
5	342

非合理的なプレイヤを選別するための基準となる戦略については，表 5.4 に示す回数 CFR を実行して生成したものを利用した．これらは対戦実験に利用した戦略のイテレーション回数（表 5.1）よりはるかに少ない．

5.3.2 実験結果

対戦結果の平均報酬は表 5.5 〜 5.7 のとおりである．なお平均標準偏差はすべて 0.1 程度である．

単純なプレイヤが常にレイズやランダムの場合はナッシュ均衡的なプレイヤより大きく報酬を得ていることがわかる．これは常にレイズやランダムが搾取しやすいだけでなく，ナッシュ均衡的な戦略では序盤からレイズが続くとフォールドを行いやすいためでもあると考えられる．

またゲームの人数，すなわち ϵ-Nash プレイヤの参加が増えるにつれ，提案手法プレイヤの平均報酬が減少していることもわかる．これは 1 人の非合理なプレイヤの行動のゲームに与える影響が少なくなるためであると考えられる．

5.3.3 考察

5.3.3.1 プレイ順序の依存性

提案手法プレイヤが単純なプレイヤから搾取したチップについては，表 5.8 のようになっている．たとえば単純なプレイヤが RAISE の場合，提案手法はプレイ順序に関わらず搾取が行えていることがわかる．

しかしながら報酬の傾向とは異なり CALL や RAND の場合は順番によらず単純プレイヤから大きく搾取できているにも関わらず，平均報酬には大きく差があることがわかる．これは第三者的なプレイヤ ϵ-Nash のゲームの参加度の違いが大きく影響しているためであると考えられる．例えば RAISE が存在する場では ϵ-Nash は強い手を警戒し，フォルドのような慎重な行動を取ることが多くなるため，ϵ-Nash のゲームへの参加度は比較的小さくなると考えられる．このことは表 5.8 の RAISE の組み合わせの平均報酬と搾取額の差がほとんど無いことからも伺える．

表 5.5: 搾取実験結果：3 人ゲーム

プレイヤの組み合わせ順序 (A, B, C)	A: 提案	B	C
(提案, CALL, ϵ-Nash)	1.00	-1.37	0.37
(提案, ϵ-Nash, CALL)	0.58	1.01	-1.59
(提案, CALL(0.5), ϵ-Nash)	-0.06	-0.19	0.25
(提案, ϵ-Nash, CALL(0.5))	-0.07	0.43	-0.36
(提案, RAISE, ϵ-Nash)	1.85	-2.01	0.15
(提案, ϵ-Nash, RAISE)	1.82	0.04	-1.87
(提案, RAISE(0.5), ϵ-Nash)	0.52	-0.73	0.21
(提案, ϵ-Nash, RAISE(0.5))	0.89	0.08	-0.97
(提案, RAISE(0.4), ϵ-Nash)	0.22	-0.62	0.39
(提案, ϵ-Nash, RAISE(0.4))	0.32	0.29	-0.61
(提案, RAND, ϵ-Nash)	1.65	-2.05	0.40
(提案, ϵ-Nash, RAND)	2.10	0.21	-2.31
(提案, RAND(0.5), ϵ-Nash)	0.59	-1.02	0.44
(提案, ϵ-Nash, RAND(0.5))	0.70	0.41	-1.10
(提案, ϵ-Nash, ϵ-Nash)	-0.64	0.35	0.28

5.3.3.2 選択した相手の妥当性

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，そのプレイヤに対してのモデル化・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を行ったため，それら相手プレイヤの行動の差が明確に現れたためであると考えられる．

相手の選択については単純プレイヤではなく，ナッシュ均衡的なプレイヤのみに注目することも有効であると考えられる．そこで選択相手を距離指標 L が最大ではなく最小のプレイヤを選択するとして対戦を行った．結果は表 5.9 の通りである．全体として L が最大のプレイヤに比べ報酬は下がり，単純なプレイヤから搾取したチップ枚数は減少している．

5.3.3.3 平均報酬の収束性

平均報酬の収束の早さについて考える．

図 5.9 は単純な相手プレイヤが CALL, CALL (0.5), ϵ-Nash の場合についての単純，ナッシュ，提案手法のプレイヤ順序の場合の提案プレイヤの平均報酬の推移である．3 つとも同じような傾向があることからも平均報酬の推移はある程度カードの偏りの影響が大きいこともわかる．

表 5.6: 搾取実験結果：4 人ゲーム

プレイヤの組み合わせ順序 (A, B, C, D)	A: 提案	B	C	D
(提案, CALL, ϵ-Nash, ϵ-Nash)	0.60	-1.60	0.45	0.55
(提案, ϵ-Nash, CALL, ϵ-Nash)	0.45	0.85	-1.74	0.44
(提案, ϵ-Nash, ϵ-Nash, CALL)	0.05	1.01	0.75	-1.81
(提案, CALL(0.5), ϵ-Nash, ϵ-Nash)	-0.18	-0.24	0.26	0.17
(提案, ϵ-Nash, CALL(0.5), ϵ-Nash)	-0.30	0.41	-0.27	0.17
(提案, ϵ-Nash, ϵ-Nash, CALL(0.5))	-0.27	0.35	0.38	-0.47
(提案, RAISE, ϵ-Nash, ϵ-Nash)	1.70	-2.06	0.21	0.15
(提案, ϵ-Nash, RAISE, ϵ-Nash)	1.87	-0.11	-1.86	0.10
(提案, ϵ-Nash, ϵ-Nash, RAISE)	1.75	0.01	0.32	-2.08
(提案, RAISE(0.5), ϵ-Nash, ϵ-Nash)	-0.05	-0.60	0.42	0.23
(提案, ϵ-Nash, RAISE(0.5), ϵ-Nash)	0.31	0.26	-0.84	0.27
(提案, ϵ-Nash, ϵ-Nash, RAISE(0.5))	0.52	0.14	0.16	-0.82
(提案, RAND, ϵ-Nash, ϵ-Nash)	1.06	-2.19	0.59	0.54
(提案, ϵ-Nash, RAND, ϵ-Nash)	1.60	0.37	-2.38	0.42
(提案, ϵ-Nash, ϵ-Nash, RAND)	1.82	0.30	0.41	-2.53
(提案, RAND(0.5), ϵ-Nash, ϵ-Nash)	-0.19	-0.84	0.60	0.44
(提案, ϵ-Nash, RAND(0.5), ϵ-Nash)	0.08	0.48	-0.95	0.39
(提案, ϵ-Nash, ϵ-Nash, RAND(0.5))	0.33	0.45	0.39	-1.17
(提案, ϵ-Nash, ϵ-Nash, ϵ-Nash)	-0.78	0.37	0.24	0.16

ルダック・ホールデムではカードのパターンが少ないため，テキサス・ホールデムより平均報酬の収束は早いと考えられる．また今回の相手である非合理なプレイヤについては戦略が単純であり，モデル化が容易であったため早期から報酬を伸ばすことが出来たと考えられる．一つの収束回数の目安として，テキサス・ホールデムのコンピューターポーカーコンペティションでは一つのプレイ順序につき 1,000 回から 3,000 回ゲームを行っていることを考えると，提案手法のモデル化での平均報酬の収束は決して早くないと判断できる．

5.3.3.4 複数の単純なプレイヤによるナッシュ均衡戦略の非合理な行動

ランダムプレイヤ（RAND）と常にレイズを行うプレイヤ（RAISE）が同時に存在する場合の対戦結果について表 5.10 に示す．ここでは搾取する提案プレイヤは参加していない．単純なプレイヤでありながら一部 ϵ-Nash より大きな報酬を得ていることがわかる．ナッシュ均衡戦略は全てのプレイヤが合理的な判断を行うことを仮定しているため，常にレイズを取るような積極的にゲームに参

表 5.7: 搾取実験結果：５人ゲーム

プレイヤの組み合わせ順序 (A, B, C, D, E)	A: 提案	B	C	D	E
(提案, CALL, ϵ-Nash, ϵ-Nash, ϵ-Nash)	-0.14	-2.12	0.75	0.79	0.72
(提案, ϵ-Nash, CALL, ϵ-Nash, ϵ-Nash)	-0.25	1.16	-2.28	0.54	0.82
(提案, ϵ-Nash, ϵ-Nash, CALL, ϵ-Nash)	-0.68	0.99	1.00	-2.22	0.91
(提案, ϵ-Nash, ϵ-Nash, ϵ-Nash, CALL)	-0.60	0.94	1.05	0.96	-2.35
(提案, RAISE, ϵ-Nash, ϵ-Nash, ϵ-Nash)	0.65	-2.69	0.68	0.56	0.79
(提案, ϵ-Nash, RAISE, ϵ-Nash, ϵ-Nash)	0.69	0.55	-2.78	0.73	0.82
(提案, ϵ-Nash, ϵ-Nash, RAISE, ϵ-Nash)	0.58	0.58	0.84	-2.72	0.71
(提案, ϵ-Nash, ϵ-Nash, ϵ-Nash, RAISE)	0.83	0.58	0.62	0.65	-2.67
(提案, RAND, ϵ-Nash, ϵ-Nash, ϵ-Nash)	0.46	-2.44	0.65	0.60	0.72
(提案, ϵ-Nash, RAND, ϵ-Nash, ϵ-Nash)	0.66	0.57	-2.58	0.64	0.71
(提案, ϵ-Nash, ϵ-Nash, RAND, ϵ-Nash)	0.91	0.52	0.50	-2.64	0.72
(提案, ϵ-Nash, ϵ-Nash, ϵ-Nash, RAND)	1.11	0.61	0.51	0.50	-2.73
(提案, ϵ-Nash, ϵ-Nash, ϵ-Nash, ϵ-Nash)	-1.22	0.41	0.36	0.24	0.22

加するプレイヤが存在する場合は早めにゲームから降りてしまい，結果として単純なプレイヤ同士の対戦になってしまうため負け越していると考えられる．

5.3.3.5 その他

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（戦略 A）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（戦略 A）を取るか，特定の相手プレイヤに着目した搾取的な戦略（戦略 B）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．また提案手法プレイヤに対し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．

表 5.8: 単純なプレイヤからの搾取したチップ

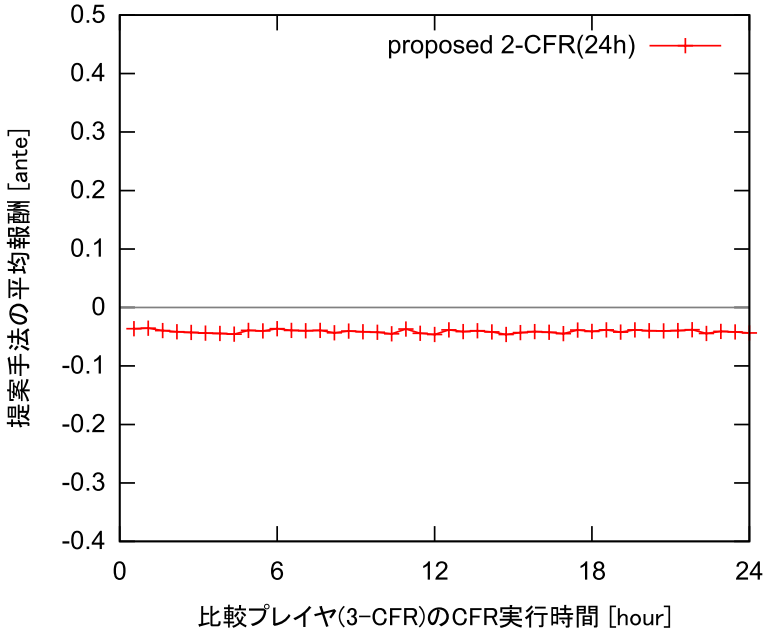
プレイヤの組み合わせ順序	提案手法の平均報酬	提案← Naïve
(提案, CALL, ϵ-Nash, ϵ-Nash)	0.60	+1.01
(提案, ϵ-Nash, CALL, ϵ-Nash)	0.45	+1.00
(提案, ϵ-Nash, ϵ-Nash, CALL)	0.05	+0.87
(提案, RAISE, ϵ-Nash, ϵ-Nash)	1.70	+1.66
(提案, ϵ-Nash, RAISE, ϵ-Nash)	1.87	+1.68
(提案, ϵ-Nash, ϵ-Nash, RAISE)	1.75	+1.72
(提案, RAND, ϵ-Nash, ϵ-Nash)	1.06	+1.47
(提案, ϵ-Nash, RAND, ϵ-Nash)	1.60	+1.74
(提案, ϵ-Nash, ϵ-Nash, RAND)	1.82	+1.90

表 5.9: 搾取実験結果：4 人ゲーム，L が最大のプレイヤを選択

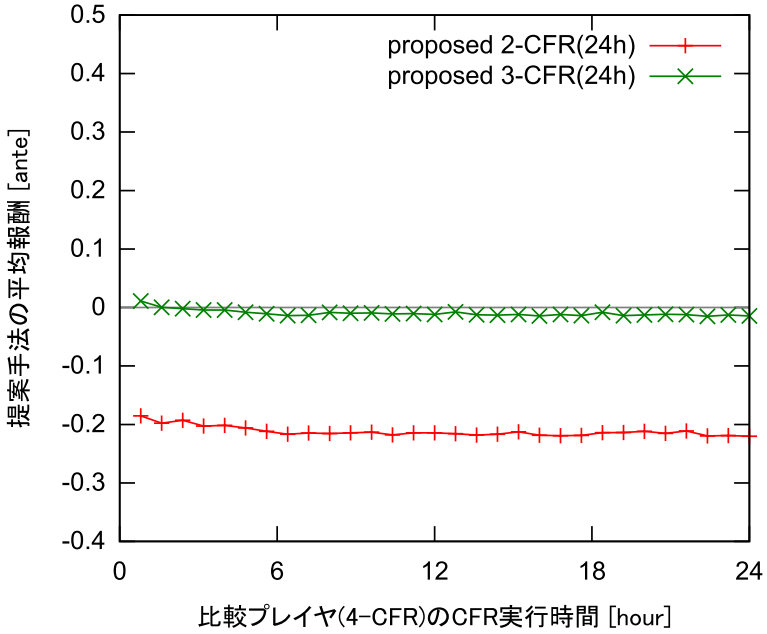
プレイヤの組み合わせ順序 (A,B,C,D)	A: 提案	B	C	D	提案← Naïve
(提案, CALL, ϵ-Nash, ϵ-Nash)	0.26	-0.87	0.48	0.14	0.33
(提案, ϵ-Nash, CALL, ϵ-Nash)	-0.23	0.95	-1.12	0.40	0.22
(提案, ϵ-Nash, ϵ-Nash, CALL)	-0.61	1.13	0.92	-1.44	0.19
(提案, RAISE, ϵ-Nash, ϵ-Nash)	1.77	-1.67	-0.01	-0.10	1.29
(提案, ϵ-Nash, RAISE, ϵ-Nash)	1.32	0.28	-1.34	-0.26	1.02
(提案, ϵ-Nash, ϵ-Nash, RAISE)	0.91	0.36	0.23	-1.50	0.81
(提案, RAND, ϵ-Nash, ϵ-Nash)	1.07	-2.09	0.58	0.44	1.33
(提案, ϵ-Nash, RAND, ϵ-Nash)	1.21	0.51	-2.18	0.46	1.44
(提案, ϵ-Nash, ϵ-Nash, RAND)	1.10	0.64	0.55	-2.29	1.48

表 5.10: 単純なプレイヤが複数存在する状態

プレイヤの組み合わせ順序	A	B	C	D	E
(RAISE, RAND, ϵ-Nash)	1.32	-1.73	0.41	–	–
(RAISE, ϵ-Nash, RAND)	1.90	0.20	-2.09	–	–
(RAISE, RAND, ϵ-Nash, ϵ-Nash)	0.66	-1.64	0.44	0.53	–
(RAISE, ϵ-Nash, RAND, ϵ-Nash)	1.12	0.36	-1.94	0.46	–
(RAISE, ϵ-Nash, ϵ-Nash, RAND)	1.47	0.43	0.43	-2.34	–
(RAISE, RAND, ϵ-Nash, ϵ-Nash, ϵ-Nash)	-0.23	-1.89	0.72	0.63	0.77
(RAISE, ϵ-Nash, RAND, ϵ-Nash, ϵ-Nash)	-0.10	0.73	-2.20	0.74	0.83
(RAISE, ϵ-Nash, ϵ-Nash, RAND, ϵ-Nash)	0.21	0.71	0.62	-2.31	0.78
(RAISE, ϵ-Nash, ϵ-Nash, ϵ-Nash, RAND)	0.41	0.78	0.58	0.74	-2.52

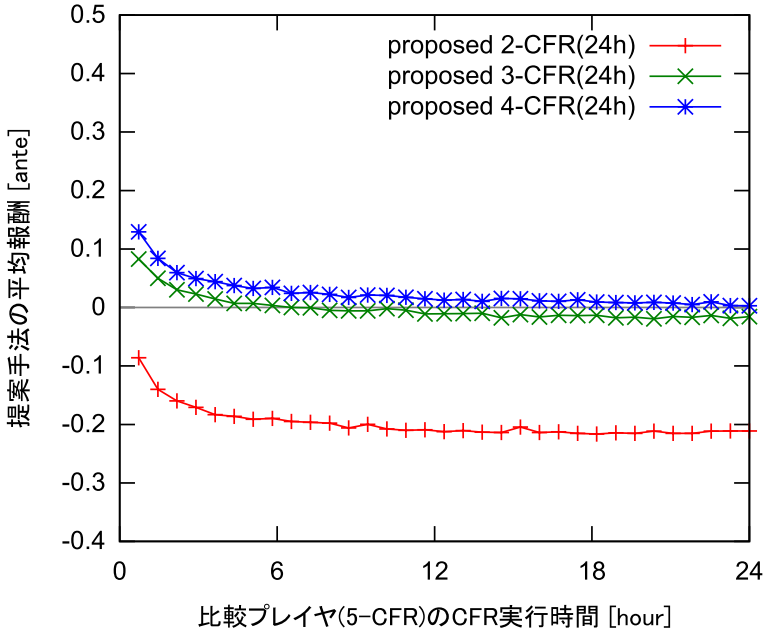


(a) 3 人ゲーム

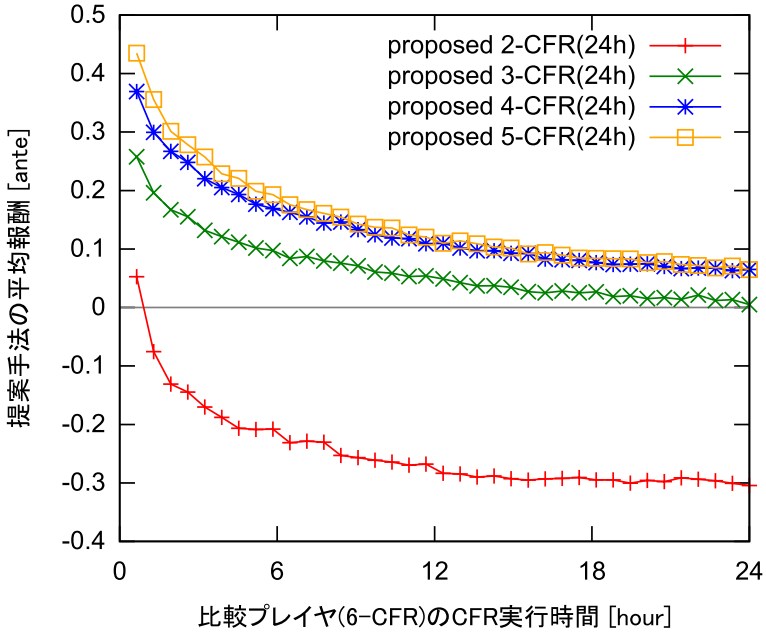


(b) 4 人ゲーム

図 5.2: 実験結果 (提案手法の CFR 実行時間を 24 時間で固定)

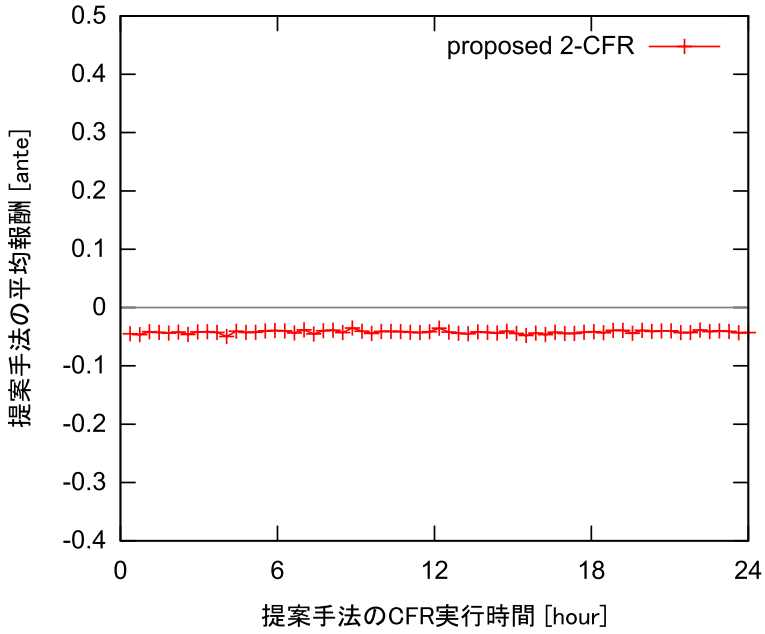


(a) 5 人ゲーム

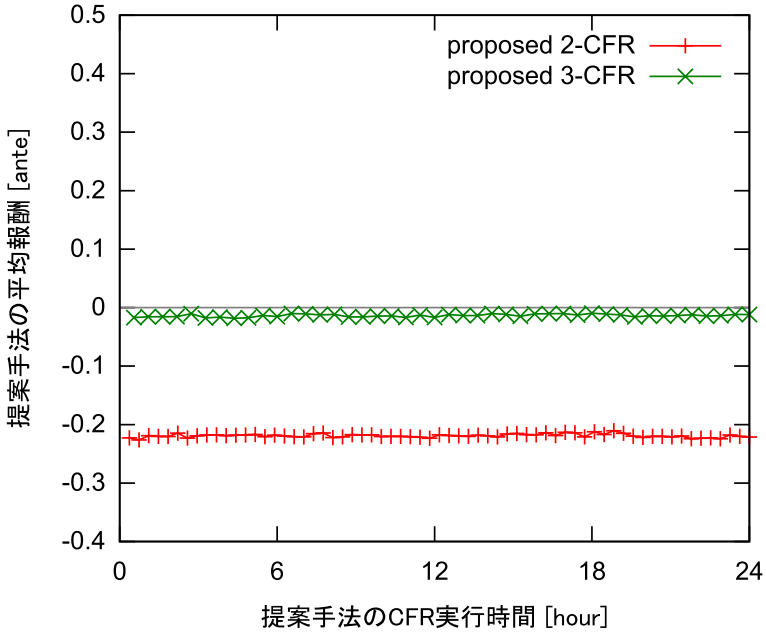


(b) 6 人ゲーム

図 5.3: 実験結果 (提案手法の CFR 実行時間を 24 時間で固定)

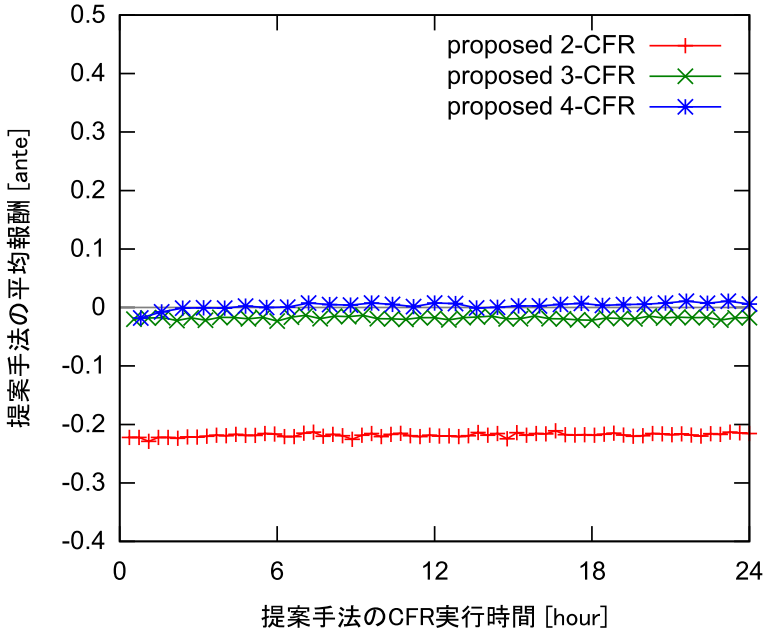


(a) 3 人ゲーム

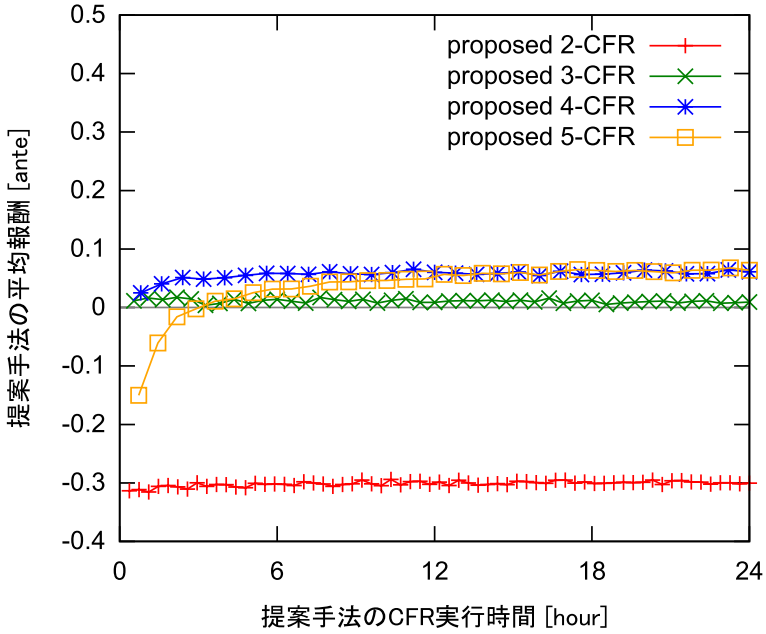


(b) 4 人ゲーム

図 5.4: 実験結果 (比較プレイヤの CFR 実行時間を 24 時間で固定)



(a) 5 人ゲーム



(b) 6 人ゲーム

図 5.5: 実験結果 (比較プレイヤの CFR 実行時間を 24 時間で固定)

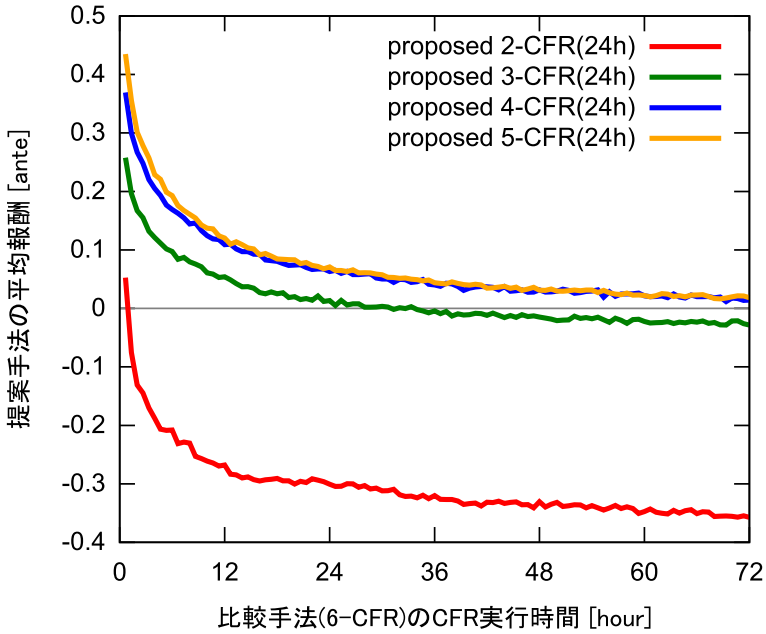


図 5.6: 6 人ゲームでの対戦結果（相手の学習時間を 72 時間まで延長）

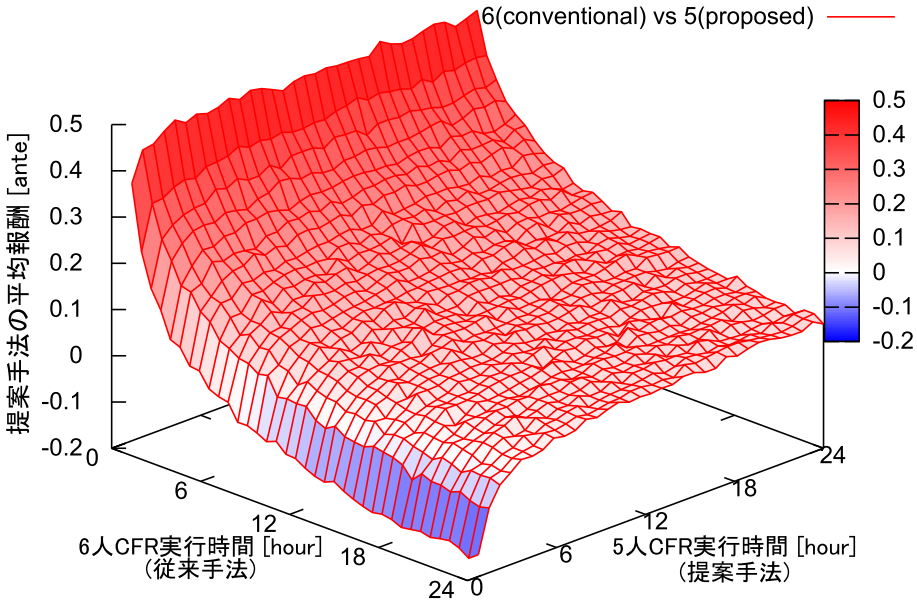


図 5.7: 6 人ゲーム対戦結果, ここでは各順序につき 50000 回の対戦を行なっている

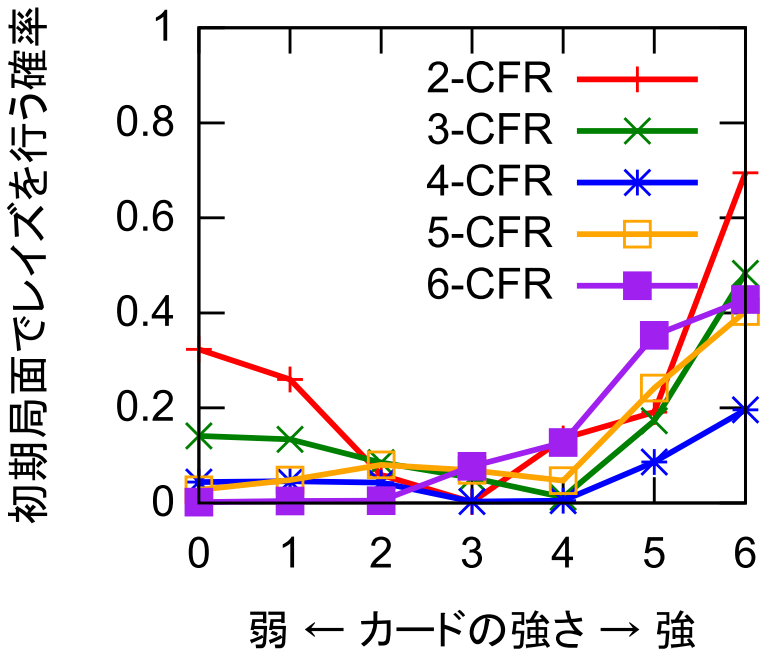


図 5.8: 初期局面でレイズを行う確率．カードは 7 種あり，ペアでなければ 0 が弱く 6 が強い

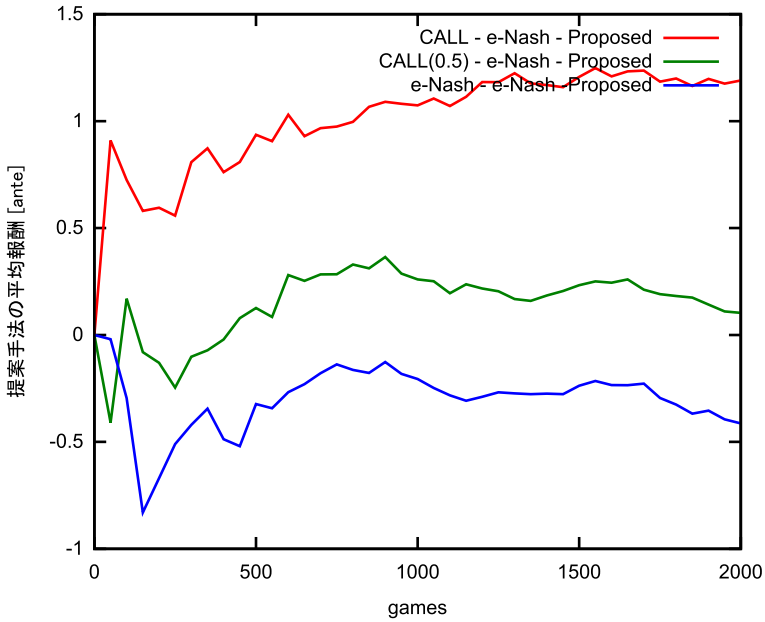


図 5.9: 平均報酬の収束性

第6章　おわりに

6.1　まとめ

本研究では不完全情報多人数ゲームの一種のポーカーにおいて，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・抽象化を用いた少人数のゲームの戦略の適用について提案し評価を行った．

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．この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．

また一方で非合理的なプレイヤが存在するような状況でそのプレイヤにのみ着目した行動を取らせることで ϵ-ナッシュ均衡戦略より大きな報酬が得られる場合があることを確認した．特に非合理的なプレイヤの影響が大きいゲーム人数の少ないゲームや，プレイヤの多くが非合理的な行動を取っている場合はナッシュ均衡戦略を取るより非合理的なプレイヤに適応するほうが良いと考えられる．

6.2　今後の課題

他のゲームへの適用可能性

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くいかない．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．

テキサスホールデムでの評価

少ない人数のナッシュ均衡戦略の適用に関しては，トイゲームだけでなく実ゲームへの適用が課題のひとつとなる．トイゲームでは 6 人ゲームでも 3 人ゲームの戦略が有効に作用したことから，テキサスホールデムでも現在の計算限界である 3 人ゲームの ϵ-ナッシュ均衡戦略が，実際にそれ以上の人数で有効に作用するかの検証を行いたい．しかしながら実ゲームでの評価自体に多大な計算量が必要となるため，評価手法自体についても再考する必要があると考えられる．

相手モデル化の改善と多人数ゲームに特化した相手モデル化

相手のモデル化に関しても今回用いたのは単純なプレイヤに有効な簡易なものであったので，ニューラルネットワークを用いた高度なモデル化などをすることで特定のプレイヤに着目しやすくなることも考えられる．また今回はプレイヤをひとりずつ減らし，そのプレイヤの行動を削減することで抽象化を行ったが，例えば複数のプレイヤの行動をまとめて一つのプレイヤで代表させるなどといった抽象化も考えられる．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．

ナッシュ均衡戦略と搾取的探索の併用

本研究では少人数のゲームでの ϵ-ナッシュ均衡戦略を適用する戦略 A と，非合理なプレイヤとの 2 人ゲームを想定した搾取的探索を行う戦略 B という，一見大きく異なる 2 つの戦略を提案したが，

- ナッシュ均衡戦略的なプレイヤしかいない場合は自身もナッシュ均衡的な戦略である戦略 A を利用
- ナッシュ均衡戦略的なプレイヤでない，非合理な行動を取っているプレイヤが 1 人ないし複数存在する場合は戦略 B を利用

というようにそれらの戦略は組み合わせて使われることが望ましい．また一方で非合理なプレイヤが存在したとしてもプレイヤ全体でみた場合の非合理性が少なくなるとやはりナッシュ均衡戦略をとることが無難であるとも考えられる．そのためにもどのような場合ならどちらの戦略を利用する，という条件を発見・設定することも重要な課題であると考えられる．

参考文献

[1] J. Rubin and I. Watson. Computer poker: A review. *Artificial Intelligence*, Vol. 175, No. 5, pp. 958–987, 2011.

[2] Michael Johanson, Nolan Bard, Neil Burch, and Michael Bowling. Finding optimal abstract strategies in extensive-form games, 2012.

[3] John Hawkin, Robert Holte, and Duane Szafron. Automated action abstraction of imperfect information extensive-form games, 2011.

[4] R. Gibson, N. Burch, M. Lanctot, and D. Szafron. Efficient monte carlo counterfactual regret minimization in games with many player actions. In *Advances in Neural Information Processing Systems 25*, pp. 1889–1897, 2012.

[5] 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.

[6] D. Koller, N. Megiddo, and B. Von Stengel. Fast algorithms for finding randomized strategies in game trees. In *Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*, pp. 750–759. ACM, 1994.

[7] 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.

[8] M. Bowling, N.A. Risk, N. Bard, D. Billings, N. Burch, J. Davidson, J. Hawkin, R. Holte, M. Johanson, M. Kan, et al. A demonstration of the polaris poker system. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*, pp. 1391–1392. International Foundation for Autonomous Agents and Multiagent Systems, 2009.

[9] J. Schaeffer, D. Billings, L. Peña, and D. Szafron. Learning to play strong poker. In *The International Conference on Machine Learning Workshop on Game Playing*, 1999.

[10] 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.

[11] 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.

[12] ポーカー侍. ポーカー教室 (カジノブックシリーズ). パンローリング, 9 2010.

[13] Michael Johanson, Kevin Waugh, Michael Bowling, and Martin Zinkevich. Accelerating best response calculation in large extensive games. In *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence - Volume Volume One*, IJCAI'11, pp. 258–265. AAAI Press, 2011.

[14] D. Billings, D. Papp, J. Schaeffer, and D. Szafron. Opponent modeling in poker. In *PROCEEDINGS OF THE NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE*, pp. 493–499. JOHN WILEY & SONS LTD, 1998.

[15] D.R. Papp. *Dealing with imperfect information in poker.* University of Alberta, 1999.

[16] D. Billings, A. Davidson, J. Schaeffer, and D. Szafron. The challenge of poker. *Artificial Intelligence*, Vol. 134, No. 1, pp. 201–240, 2002.

[17] R. Coulom. Efficient selectivity and backup operators in monte-carlo tree search. *Computers and Games*, pp. 72–83, 2007.

[18] D. Billings, J. Schaeffer, D. Szafron, et al. Using probabilistic knowledge and simulation to play poker. In *Proceedings of the National Conference on Artificial Intelligence*, pp. 697–703. JOHN WILEY & SONS LTD, 1999.

[19] D. Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, Vol. 6, No. 1, pp. 1–8, 1956.

[20] M.B. Johanson. 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.

[21] J. Hawkin, R. Holte, and D. Szafron. Automated action abstraction of imperfect information extensive-form games. In *National Conference on Artificial Intelligence (AAAI)*, 2011.

[22] K. Waugh, N. Bard, and M. Bowling. Strategy grafting in extensive games. *Advances in Neural Information Processing Systems*, Vol. 22, pp. 2026–2034, 2009.

[23] A. Davidson, D. Billings, J. Schaeffer, and D. Szafron. Improved opponent modeling in poker. International Conference on Artificial Intelligence, ICAI ' 00, pp. 1467–1473, 2000.

[24] 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.

[25] C. Luckhardt and K.B. Irani. An algorithmic solution of n-person games. In *Proceedings of the 5th National Conference on Artificial Intelligence (AAAI)*, Vol. 1, pp. 158–162, 1986.

[26] N.R. Sturtevant and R.E. Korf. On pruning techniques for multi-player games. In *Proceedings of the National Conference on Artificial Intelligence*, pp. 201–208. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2000.

[27] N. Sturtevant. A comparison of algorithms for multi-player games. *Computers and Games*, pp. 108–122, 2003.

[28] M.P.D. Schadd and M.H.M. Winands. Best reply search for multiplayer games. *Computational Intelligence and AI in Games, IEEE Transactions on*, Vol. 3, No. 1, pp. 57–66, 2011.

[29] B.W. Ballard. The*-minimax search procedure for trees containing chance nodes*. *Artificial Intelligence*, Vol. 21, No. 3, pp. 327–350, 1983.

[30] M. Lanctot, K. Waugh, M. Zinkevich, and M. Bowling. Monte carlo sampling for regret minimization in extensive games. *Advances in Neural Information Processing Systems*, Vol. 22, pp. 1078–1086, 2009.

発表文献

査読付き会議論文

1. 古居 敬大, 三輪 誠, 近山 隆. 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. 第 16 回ゲームプログラミングワークショップ, 2011.11. **(ゲームプログラミングワークショップ研究奨励賞)**

2. 古居 敬大, 浦 晃, 三輪 誠, 鶴岡 慶雅, 近山 隆. 相手の抽象化による多人数ポーカーでの戦略の決定. 第 17 回ゲームプログラミングワークショップ, 2012.11. **(ゲームプログラミングワークショップ優秀論文賞)**

謝辞

本論文を書くにあたって多くの方々にお世話になりました.

指導教員である近山隆教授には自身の研究内容にとどまらない研究や考え方についてのご意見・ご指摘や，発表等における様々なアドバイスをいただきました.

鶴岡慶雅准教授には AI 勉強会などで多くの鋭いご意見をいただき，自分の研究をより良いものにすることが出来ました．また将棋プログラム激指にも大きな刺激を受けました．今後趣味として将棋プログラムを自作していけたらと思っています.

近山研究室 OB のマンチェスター大学の三輪誠先輩や，博士課程の浦晃先輩には研究の初期から研究テーマに関して相談に乗っていただきました．大学院から始めた学部時代とは全く異なる研究分野について，円滑に研究を進めることが出来たのは先輩方のおかげであったと思います.

学部学生の頃からの同期である修士課程の関栄二君とは日頃から活発な議論を行い，互いの研究について理解を深め合うことが出来ました．また，ここには書ききれませんが，他の近山・鶴岡研，田浦研究室の多くの皆様にも支えられて研究生活を送ることが出来ました．この場をお借りして厚くお礼申し上げます.

平成 25 年 2 月 6 日