ゲーム局面における特徴量の重要度の定量的評価

藤井 慧[1,a)]

概要：これまでゲームの強い戦略を求めることを目的とした多くの研究が成功を収めてきたが，その戦略を人間に解釈可能な形で説明する研究は進んでおらず，近年注目を集めつつある．本研究では，これまでCFR 等のアルゴリズムを大きなゲームに適用して強い戦略を計算するために用いられてきた state-space abstraction の手法を用いて，ゲーム局面における特徴量やその組み合わせの重要度を定量的に評価する手法を提案．また，Goofspiel を題材として実験を行い，この手法による分析でゲームの特徴量やそれを用いた意思決定に関する知見が得られることを示した．

Evaluation of Feature Importance in Game States

Fujii Satoru[1,a)]

Abstract: There has been a lot of successful research about finding the strong strategy in games, but their methods can't explain their strategies to humans. State-space abstraction is a technique widely applied to approximate optimal policies in large games. In this paper, however, we propose a method to evaluate the importance of features in game states via state-space abstraction. We present experimental results which shows this analysis allows us to understand the game and improve our decisions.

1. はじめに

ある程度以上の大きさを持つ不完全情報ゲームについてナッシュ均衡を完全に求めるのは，コンピュータを用いても困難であることが多い．そのため，これまでに不完全情報ゲームのナッシュ均衡を近似的に求めるための様々なアルゴリズムが研究，提案されてきた．その中でも，ゲームの遷移規則が与えられている設定においては，Counterfactual Regret Minimization (CFR) [1] やその派生アルゴリズムが良い成果を挙げてきた．

また，これらの手法を大きなゲームに適用するために，state-space abstraction[2][3] などを用いてゲームの抽象化を行う研究や，それらの抽象化の精度を評価・比較する研究がこれまで行われてきた．

一方で，こういった強い戦略を計算することを目的としたアルゴリズムは，その結果を人間に解釈可能な形で表現することを前提としておらず，意思決定の根拠を人間に利用可能な形で説明することができない．そこで，こうした研究とは別に，人間の意思決定の精度を向上させたり，ゲームの理解度を深めたりすることを目的とした研究も行われるようになってきている．

人間は局面のいくつかの特徴量に着目して意思決定を行っていると考えられる．例えば得点差を見て，相手に差を付けられていればハイリスク・ハイリターンな行動で一発逆転を狙ったり，十分相手を上回っていればローリターンでも安定した行動を選んだりといった具合に，局面の持ついくつかの特徴量を手掛かりに，ある種パターン認識的に最適な行動を推測するのである．

こういったゲーム局面の特徴量やその組の重要度を定量的に評価できれば，それはそれ自体ゲームの性質の分析として興味深いだけでなく，着目すべき情報を示唆して人間の意思決定に寄与することも期待できる．

これまで state-space abstraction は，大きなゲームに対して CFR 等の手法を適用し，ナッシュ均衡により近い強い戦略を計算するために用いられてきたが，本研究では

[1] 京都大学大学院 人間・環境学研究科
Graduate School of Human and Environmental Studies, Kyoto University
[a)] fujii.satoru.75c@st.kyoto-u.ac.jp

state-space abstraction を利用して，不完全情報ゲームの局面における特定の特徴量やそれらの組み合わせがゲーム中の意思決定に関してどの程度の重要度を持っているかを定量的に評価する手法を提案する．また，Goofspiel を題材とした実験を行い，この手法を通じてゲームについての興味深い知見が得られることを示す．

2. 関連研究

2.1 展開型ゲームとナッシュ均衡

以下の 5 要素の組によるゲームの表現形式を，展開型ゲーム [4] と呼ぶ．

(1) ゲーム木 K：ゲームの局面の遷移規則を表現した根付き有限有向木．終端ノードはゲーム終了を表し，終端ノード全体を Z とかく．終端ノード以外のノードを手番といい，手番全体を X とかく．手番 x から終端ノード方向に伸びる辺を選択肢といい，x における選択肢の集合を $A(x)$ とかく．ゲーム木のノードのことを履歴（history）とよぶ．

(2) プレイヤー分割 P：X の分割 $P = (P_0, P_1, \cdots, P_n)$．ここで，$n$ はプレイヤー数で，集合 P_i はプレイヤー i の手番全体を表現する．但し，P_0 は偶然手番，つまりいかなるプレイヤーの意思とも無関係な確率的な局面の分岐を表現する．

(3) 偶然手番の確率分布族 p：各偶然手番 $x \in P_0$ において各選択肢に付与される確率の分布の族．

(4) 情報分割 U：プレイヤー分割 P の細分割 $U = (U_0, U_1, \cdots, U_n)$．ここで U_i は P_i を分割する部分集合族．U_i に属する集合をプレイヤー i の情報集合という．$I \in U_i$ について，ゲームが手番 $x \in I$ に到達したとき，手番プレイヤー i は現在の手番が I の元であることは認識できるが，具体的にその中のどの元に到達したかを区別できない．但し，U_0 に属する全ての情報集合はただ 1 つの偶然手番からなるとする．情報集合 I における選択肢の集合を $A(I)$ とかく．

(5) 利得関数 u：ゲーム木の終端ノードそれぞれに対して，各プレイヤーの利得を表す n 成分ベクトルを対応させる関数．全ての終端ノードで全プレイヤーの利得の和が 0 となるとき，ゲームは零和であるという．

全てのプレイヤーの全ての情報集合がただ 1 つの手番からなるとき，その展開型ゲームは完全情報ゲームであるといい，そうでない場合は不完全情報ゲームであるという．

プレイヤー i の各情報集合 $I \in U_i$ に対して，各選択肢 $a \in A(I)$ の選択確率（行動戦略という）を対応させる関数 σ_i をプレイヤー i の行動戦略という．プレイヤー i の行動戦略全体の集合を Σ_i とかく．プレイヤー全体の行動戦略の組 σ を戦略プロファイルという．また，戦略プロファイル σ からプレイヤー i の戦略 σ_i を除いたものを σ_{-i} とかく．

戦略プロファイル σ に従って各プレイヤーが行動したときの履歴 h，情報集合 I への到達確率をそれぞれ $\pi^\sigma(h)$，$\pi^\sigma(I)$ とかき，σ のもとでのプレイヤー i の期待利得を，

$$u_i(\sigma) = \sum_{h \in Z} \pi^\sigma(h) u_i(h)$$

と定義する．

戦略プロファイル $\sigma^* = (\sigma_1^*, \cdots, \sigma_n^*)$ において，すべてのプレイヤー $i = 1, \cdots, n$ のすべてのとりうる戦略 σ_i に対して

$$u_i(\sigma^*) \geq u_i(\sigma_i, \sigma_{-i}^*)$$

が成立するとき，σ^* はナッシュ均衡であるという．このような戦略プロファイルは，必ずしも一意には定まらない．

2.2 可搾取量

プレイヤー i の，他のプレイヤーの戦略 σ_{-i} に対する最適応答 $b_i(\sigma_{-i})$ を，

$$b_i(\sigma_{-i}) = \max_{\sigma_i' \in \Sigma_i} u_i(\sigma_i', \sigma_{-i},)$$

で定義する．また，戦略プロファイル σ に対する可搾取量 $\epsilon(\sigma)$ を，

$$\epsilon(\sigma) = \frac{\sum_{i \in N} b_i(\sigma_{-i})}{n}$$

で定義する（但し，N はプレイヤー全体の集合）[5]．これは，与えられた戦略プロファイルがナッシュ均衡にどの程度近いか評価するための指標として有用である[*1]．

2.3 Counterfactual Regret Minimization (CFR)

Counterfactual Regret Minimization (CFR)[1] とは，不完全情報ゲームのナッシュ均衡の 1 つを近似的に求めるためのアルゴリズムである．

戦略プロファイル σ のもとでの履歴 $h \notin Z$ に対するプレイヤー i の counterfactual value を，

$$v_i(\sigma, h) = \sum_{z \in Z, h \sqsubset z} \pi_{-i}^\sigma(h) \pi^\sigma(h, z) u_i(z)$$

で定義する．但し，$h \sqsubset z$ は h から z に到達可能であることを，$\pi_{-i}^\sigma(h)$ はプレイヤー i が h へ向けて行動するように σ を変更したときの h への到達確率を，$\pi^\sigma(h, z)$ は σ のもとでの h から z への到達確率をそれぞれ表している．

また，σ において，履歴 h，情報集合 I で，手番プレイヤー i が選択肢 a を選ばないことに対する counterfactual regret をそれぞれ

$$r^\sigma(h, a) = v_i(\sigma_{I \to a}, h) - v_i(\sigma, h)$$

$$r^\sigma(I, a) = \sum_{h \in I} r^\sigma(h, a)$$

[*1] ナッシュ均衡における可搾取量は 0 になる．

と定義する．但し，$\sigma_{I\to a}$ は σ における I での a の選択確率を 1 に変更した戦略プロファイルを表す．

CFR では，各タイムステップごとに戦略プロファイルを更新していく．

アルゴリズムのタイムステップ T における累積 counterfactual regret を

$$R^T(I,a)=\sum_{t=1}^{T} r^{\sigma_t}(I,a)$$

と定義する．また，$R^T_+(I,a)=\max(R^T(I,a),0)$ とする．

ここで，CFR の戦略プロファイルの更新は，

$$\sigma^{T+1}(I,a)=\begin{cases}\dfrac{R^T_+(I,a)}{\sum_{a\in A(I)} R^T_+(I,a)} & \text{if 分母 }>0\\ \dfrac{1}{|A(I)|} & \text{otherwise}\end{cases}$$

によって行われる．ゲームが零和で，プレイヤー数が 2 のとき，このように更新していった戦略プロファイル $\sigma_1,\cdots,\sigma_T$ の平均が Nash 均衡戦略のひとつに収束していくことが知られている．

累積 regret の更新を両方のプレイヤーで同時に行うのではなく，片方のプレイヤーのみに限定して交互に入れ替える alternating update と呼ばれる手法も存在し，経験的にはより速く Nash 均衡に収束することが知られている [6][7]．このとき，累積 regret を更新するプレイヤーを traverser という．

2.4 Monte Carlo CFR (MCCFR)

CFR では counterfactual regret の計算のために各タイムステップごとにゲーム木を全て探索する必要があり，これはゲーム木が大きい場合には困難である．Monte Carlo CFR (MCCFR) [8] は，ゲーム木の探索を部分的にしか行わないことによってこの問題を解決する CFR の発展的手法の 1 つである．

$Q=(Q_1,\cdots,Q_r)$ を Z の部分集合族とする．但し，$\bigcup_{i=1,\cdots,r} Q_i=Z$ となるようにする．これらの部分集合それぞれを block とよぶ．各タイムステップにおける木の横断を全ての終端ノードに対してではなく，ただ 1 つの block に含まれる終端ノードに対してのみ行うことを考える．

このとき，各タイムステップでそれぞれの block を参照する確率を $q=(q_1,\cdots,q_r)$ とする．但し，q_i は Q_i を参照する確率である．このとき，終端ノード $z\in Z$ が各タイムステップで参照される確率は $\sum_{j:z\in Q_j} q_j$ であり，これを単に $q(z)$ とかく．

ブロック Q_j を参照するときの sampled counterfactual value を，counterfactual value と同様の設定で

$$\tilde{v}_i(\sigma,I|j)=\sum_{h\in I}\sum_{z\in Q_j, h\sqsubset z}\frac{1}{q(z)}\pi^{\sigma}_{-i}(h)\pi^{\sigma}(h,z)u_i(z)$$

と定義する．sampled counterfactual value の期待値は counterfactual value と一致し，MCCFR ではこれを用いて計算した sampled counterfactual regret を用いる．

とくに，alternating update を用い，Z を traverser でないプレイヤーの純戦略と偶然手番の結果ごとに分割したものを Q とし，q を σ_{-i} とに応じた到達確率で定めたものを External-Sampling MCCFR と呼ぶ．これは，ナッシュ均衡戦略に確率的に収束していくことが理論的に示されている．

2.5 ゲームの抽象化

CFR や MCCFR では全ての情報集合に対して counterfactual regret と平均戦略を格納するが，これは大きいゲームを対象とする場合は困難である．プレイヤーが複数の（似ている）情報集合を区別できず，それらの情報集合の間で同じ局所戦略を持つ状態で学習することにより使用メモリを削減する state-space abstraction[2][3] はこの問題を解決する手法の 1 つである．

2 人のプレイヤーのうち片方にはこのような抽象化を行わない場合，state-space abstraction を用いた学習による戦略が，抽象化の範囲で実現可能なもののうち，元のゲームのナッシュ均衡に可搾取量の意味で最も近い戦略に収束していくことが示されている [9]．

具体的な state-space abstraction の方法としては，k 平均法を用いるもの [10] やポテンシャルを考慮するもの [11] などがこれまでに提案されており，これらの手法のパフォーマンスを，可搾取量や特定の相手に対する期待利得を用いて評価する研究も行われてきた [12][3]．特定の相手に対する期待利得の大きさは必ずしも戦略のナッシュ均衡への近さを意味しないが，全ての情報を利用できるプレイヤーに対してどの程度負けるかという値は可搾取量よりも直感的に理解しやすいことも考えられ，計算時間が短い利点もある．

また，ここまでに述べた収束性の証明は全て，各プレイヤーが自身の行動の履歴を記憶している完全記憶性を前提としているが，完全記憶性を持たない抽象化を行った場合でも良い結果が得られることが経験的に確かめられている [13]．

3. 提案手法

分析対象とするプレイヤー i を選び，そのプレイヤーの各情報集合 $I\in U_i$ から意思決定を説明する特徴量を与える関数の組 $V=(V_0,V_1,...,V_m)$ を設計する．特徴量の形は離散的であれば何でもよく，$V_0(I)$=$A(I)$ とする．

V_0 を含む V の部分集合 S を選び，$S(I)=(S_0(I),S_1(I),...,S_{m'}(I))$ の全ての値が等しい複数の $I\in U_i$ を同一視することによる state-space abstraction を行って学習を行う．学習アルゴリズムとしては CFR や MCCFR

を用いる．これは，プレイヤー i が S で得られる特徴量のみを手掛かりに学習している事を意味する．

こうして得た各 S による学習結果を可搾取量によって評価・比較することで，それぞれの特徴量やその組み合わせが意思決定に及ぼす影響の大きさを計測する．また，学習結果における，抽象化を行っていないプレイヤーに対する期待利得を評価に用いることもできる．

4. 実験

4.1 Goofspiel

本研究では，実験の題材として Goofspiel[14] を用いた．Goofspiel は 2 人以上のプレイヤーによる不完全情報ゲームで，ルールにはいくつかのバリアントが存在するが，本研究で用いたものは以下の通りである．

2 人のプレイヤーで行う．両方のプレイヤーは，1 から N までの数が書かれたカードを 1 枚ずつ，計 N 枚の手札を持った状態でゲームを開始する．また，山札としてプレイヤーの手札と同様に 1〜N までの N 枚のカードを用意し，シャッフルする．

ゲームは N ラウンドで構成される．各ラウンドではまず山札から 1 枚のカードが場の中央に公開される．両方のプレイヤーは公開されたカードの数を見たあと，自分の残り手札の中から任意のカード 1 枚を選択して，同時に提示する．このとき，提示したカードの数が大きかったプレイヤーは中央のカードに書かれた数の分だけ得点を得る．なお，提示したカードの数が等しい場合にはお互いに得点は得られない．中央に公開されたカードとお互いのプレイヤーが提示したカードはゲームから除外される．全てのラウンドの終了時，得点の合計が多かったプレイヤーが勝利する．

本研究では $N = 5$ として実験を行った．なお，Goofspiel のルールは同時手番を繰り返す形になっているが，まず一方のプレイヤーがカードを提示し，次にそのカードの数を知ることなくもう一方のプレイヤーがカードを提示するという形にすることで，ゲームの性質を変えることなく展開型ゲームとして表現できる．

4.2 設定

片方のプレイヤーのみを抽象化して学習し，アルゴリズムとして External Sampling MCCFR[8] を，学習環境として OpenSpiel[15] を用いた．Goofspiel は対称なゲームなのでもう片方のプレイヤーに対して再度学習する必要はなく，また，お互いが全ての情報を利用できるときの Nash 均衡における期待利得は 0 である．

Goofspiel の情報集合において，合法手決定のために必要となる自分の手札の情報を与える V_0 の他に，ゲームの履歴から得られる情報のうち意思決定に本質的に関与する特徴量として，

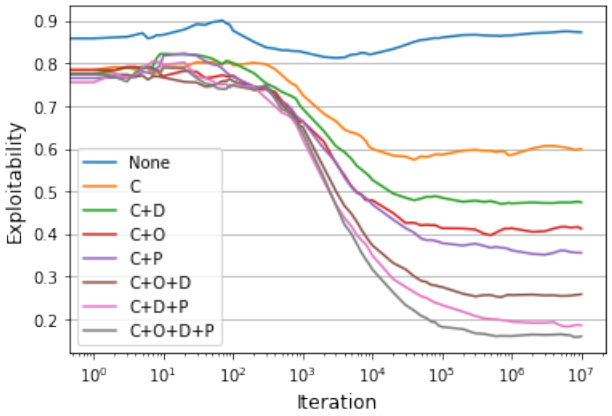

図 1 可搾取量 (1)
Fig. 1 Exploitability (1)

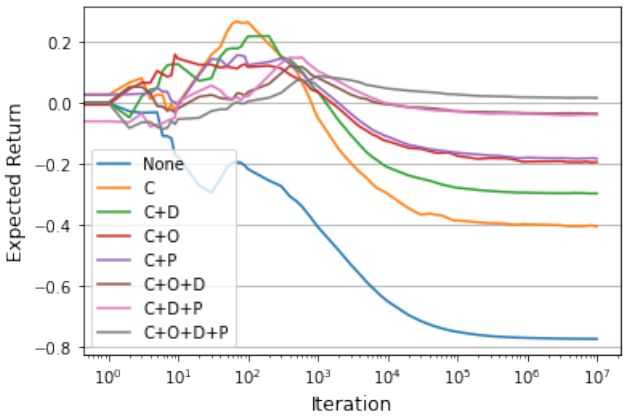

図 2 抽象化を行ったプレイヤーの期待利得 (1)
Fig. 2 Expected return of the player with abstraction (1)

- 現在中央に公開されているカード（C）
- 相手の残り手札（O）
- 山札に残っているカード（D）
- 自分の得点から相手の得点を引いた値（P）

の 4 つを加えて V を設計した．これらは十分な情報集合の表現力を持っている．[*2]

以下，これらの特徴量のうち，例えば C と O を与える関数を S として選んだ場合の学習結果を C+O といったように表記する．

4.3 結果

上で定めた V のいくつかの部分集合を用いたときの学習結果を，**図 1** 及び**図 2** に示した．利用可能な情報が多いほど可搾取量・期待利得が 0 に近付いていることに注意されたい．これらの結果から，相手の残りカードを考慮しない戦略（C+D+P），得点差を（ほぼ）考慮せずその時点からの得点に集中する戦略（C+O+D）でも情報を全て利用する戦略とほぼ遜色ないパフォーマンスが発揮できるこ

[*2] $V(I)$ と I は一対一対応しないが，$V(I)$ の全ての値が等しい複数の情報集合は本質的に同じゲーム状態を表している．

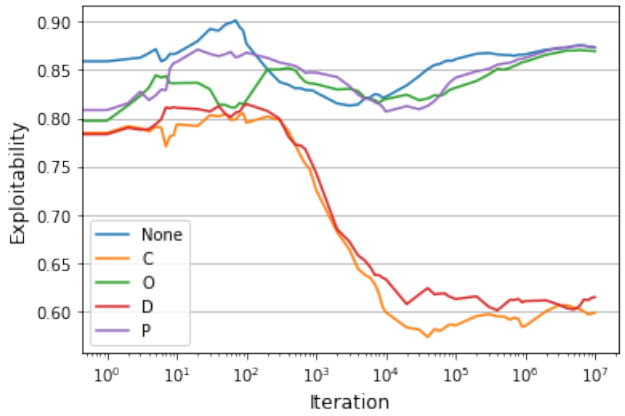

図 3 可搾取量 (2)
Fig. 3 Exploitability (2)

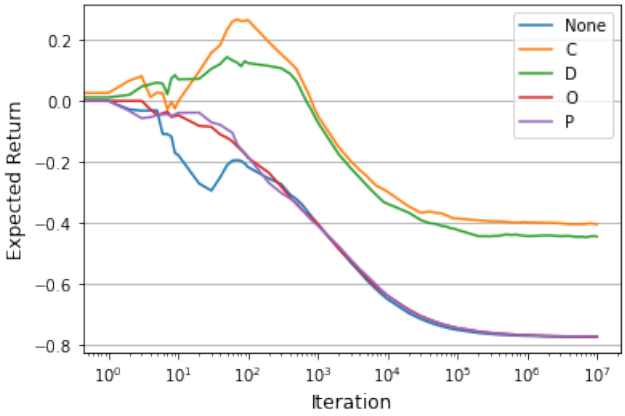

図 4 抽象化を行ったプレイヤーの期待利得 (2)
Fig. 4 Expected return of the player with abstraction (2)

となどが分かる．また，C+O+D+P は抽象化を行わないプレイヤーの期待利得を上回っており，特徴量を用いた抽象化によって学習が効率化されたことを示している．

また，それぞれの特徴量を単独で利用する抽象化を用いて学習した結果を，**図 3** 及び**図 4** に示した．この結果からは，C や D が単独でもかなりの重要度を持っていて，C や D だけを見て手札から提示するカードを決める戦略でも無情報の戦略よりかなり良いパフォーマンスが得られることが分かる．これは，場に出ているカードが大きければ優先して取ろうとしたり，山札に大きなカードが残っていれば手札の大きなカードを温存しようとしたりするなどの判断の重要性を示唆している．

逆に，O や P はそれ単独ではあまり意味を持たないことが分かる．他の結果と合わせれば，O や P は決して無意味な特徴量ではなく，C や D と組み合わせたときにのみ意味を持つことが分かる．このことは直感的にも，現在中央に公開されているカードや山札に残っているカードについて情報が得られてはじめて，相手の手札と自分の残り手札を比較してどこで勝ってどこで負けるかを計画することが可能になったり，得点差を見て相手に勝つために具体的に取らなければいけないカードの組み合わせを知ることができたりするなどと解釈できる．このように，特徴量の重要性は，利用可能な他の特徴量と相互に結びついていることが分かる．

特徴量間の相関関係を用いて，ある特徴量から他の特徴量をある程度推測可能な場合があることも注意したい．例えば，C と D はお互いからお互いを推測可能な関係にあると考えられる[*3]．C と D は単独では O や P よりも良いパフォーマンスを発揮しているが，この 2 つを組み合わせた C+D よりも，C+O や C+P の方が良いパフォーマンスを発揮していることから，相関関係の強い特徴量を組み合わせてもパフォーマンスがそれほど向上しないことが推察できる．

なお，S として他の部分集合を用いたときの結果は，**図 5** 及び**図 6** に示した．

5. おわりに

本研究では，ゲームの戦略や性質を人間に利用可能な形で分析することを目的として，state-space abstraction を用いて局面の特徴量の重要度を評価する手法を提案し，Goofspiel を用いた実験で，提案手法がゲーム中の意思決定における重要な知見を与えてくれることを示した．

提案手法の限界として，合法手の決定に必要な特徴量だけで情報集合をほとんど表現してしまうようなゲームや，特徴量の組が上手く作れないようなゲームに対しては適用が困難であることが挙げられる．そういったゲームでは，そもそも特徴量に着目して分析を行うこと自体が難しいかもしれない．

また，1 人のプレイヤーでしか抽象化を行わないため，大きすぎるゲームに対しては適用が難しい．こういった場合には，例えば学習アルゴリズムとして Deep CFR[7] を採用し，ニューラルネットの入力値に $S(I)$ を用いることで同様の分析ができると考えられるが，これは今後の課題としたい．

本手法による分析は，単にゲームの性質についての理解を深めてくれるだけでなく，判断の手掛かりとすべき着眼点を示すことで人間の意思決定の精度向上に役立てられると期待でき，有意義であると考えている．

謝辞 本稿執筆にあたって有益な助言を頂いた立木秀樹教授に，この場を借りて謝意を記します．また，コメントを頂いた研究室の方々や査読者の方々にも謝意を記します．

参考文献

[1] 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 (2007).

[*3] 例えば，最初の手番ではこの 2 つの特徴量は本質的には同じ

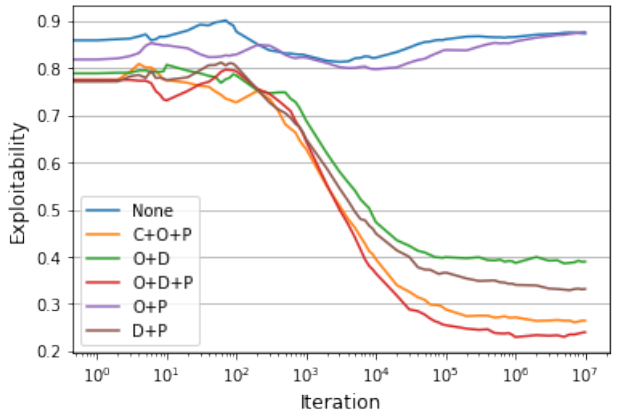

図 5 可搾取量 (3)
Fig. 5 Exploitability (3)

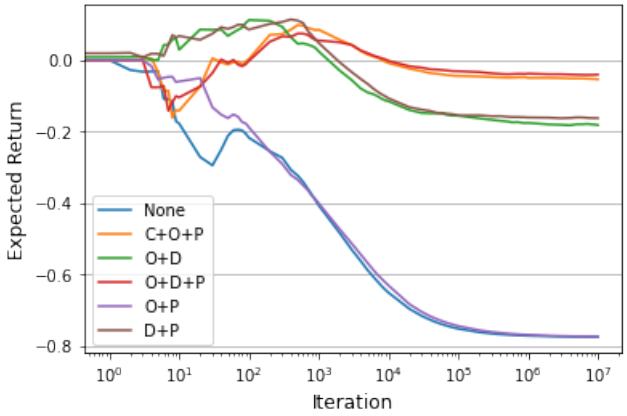

図 6 抽象化を行ったプレイヤーの期待利得 (3)
Fig. 6 Expected return of the player with abstraction (3)

[2] Shi, J. and Littman, M. L.: Abstraction methods for game theoretic poker, *International Conference on Computers and Games*, Springer, pp. 333–345 (2000).

[3] Johanson, M., Burch, N., Valenzano, R. and Bowling, M.: Evaluating state-space abstractions in extensive-form games, *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, pp. 271–278 (2013).

[4] 岡田章：ゲーム理論，有斐閣 (2011).

[5] Johanson, M., Bard, N., Burch, N. and Bowling, M.: Finding optimal abstract strategies in extensive-form games, *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 26, No. 1 (2012).

[6] Tammelin, O.: Solving large imperfect information games using CFR+, *arXiv preprint arXiv:1407.5042* (2014).

[7] Brown, N., Lerer, A., Gross, S. and Sandholm, T.: Deep counterfactual regret minimization, *International conference on machine learning*, PMLR, pp. 793–802 (2019).

[8] 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).

[9] Waugh, K., Schnizlein, D., Bowling, M. H. and Szafron, D.: Abstraction pathologies in extensive games., *AAMAS (2)*, pp. 781–788 (2009).

[10] Gilpin, A. and Sandholm, T.: Better automated abstraction techniques for imperfect information games, with application to Texas Hold'em poker, *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, pp. 1–8 (2007).

[11] Gilpin, A., Sandholm, T. and Sørensen, T. B.: Potential-aware automated abstraction of sequential games, and holistic equilibrium analysis of Texas Hold'em poker, *Proceedings of the National Conference on Artificial Intelligence*, Vol. 22, No. 1, Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, p. 50 (2007).

[12] Gilpin, A. and Sandholm, T.: Expectation-Based Versus Potential-Aware Automated Abstraction in Imperfect Information Games: An Experimental Comparison Using Poker., *AAAI*, pp. 1454–1457 (2008).

[13] Waugh, K., Zinkevich, M., Johanson, M., Kan, M., Schnizlein, D. and Bowling, M.: A practical use of imperfect recall, *Eighth symposium on abstraction, reformulation, and approximation* (2009).

[14] Ross, S. M.: Goofspiel—the game of pure strategy, *Journal of Applied Probability*, Vol. 8, No. 3, pp. 621–625 (1971).

[15] Lanctot, M., Lockhart, E., Lespiau, J.-B., Zambaldi, V., Upadhyay, S., Pérolat, J., Srinivasan, S., Timbers, F., Tuyls, K., Omidshafiei, S., Hennes, D., Morrill, D., Muller, P., Ewalds, T., Faulkner, R., Kramár, J., Vylder, B. D., Saeta, B., Bradbury, J., Ding, D., Borgeaud, S., Lai, M., Schrittwieser, J., Anthony, T., Hughes, E., Danihelka, I. and Ryan-Davis, J.: OpenSpiel: A Framework for Reinforcement Learning in Games, *CoRR*, Vol. abs/1908.09453 (online), available from ⟨http://arxiv.org/abs/1908.09453⟩ (2019).