解釈可能なポーカー AI による弱点モデル化と相手適応

大端 俊輝[1,a)]　鶴岡 慶雅[1,b)]

概要：ポーカーの代表的な種目であるノーリミットテキサスホールデム（NLH）の学習では，ゲーム理論上の均衡戦略（GTO; Game Theory Optimal）を数値計算するソルバが広く用いられている．しかしその出力は「どの手札でどの行動をどれだけの割合で選ぶか」という頻度表であり，何をすべきかは示しても，なぜそうするのかは説明しない．さらに，2 人対戦のゼロサムゲームにおける GTO は，相手によらず長期的に搾取されないことを目指す静的な戦略であるため，目の前の相手が持つ悪癖（リーク）を積極的に突く戦略とは本質的に異なる．本研究では，著者らが強い人間プレイヤーの段階的思考と GTO の構造的基準を手作業で明示的ルールに体系化したルールベース AI「advance0」を土台に，(1) 参加する手札の広さ（ルース／タイト）と攻撃性の高低（アグレッシブ／パッシブ）の 2 軸で典型的な悪癖を再現した弱い AI，および (2) 外部から与えられた相手型に応じて対策を切り替えるエクスプロイト Bot を構築した．評価実験では，advance0 は GTO と高い一致度を示し，エクスプロイト Bot は弱い AI の 4 型のうち 3 型に対して，相手の型を仮定しない基準戦略（base）を有意に上回った．さらに，相手のタイプごとに有効な対策が異なることを確認した．これらの結果は，解釈可能なルールによって GTO 近似のみならず，相手適応的な搾取戦略までを人間が追跡可能な形で構成できることを示している．

キーワード：ポーカー AI, 不完全情報ゲーム, 解釈可能な AI, GTO（Game Theory Optimal）, 相手モデリング, 搾取戦略

Weakness Modeling and Opponent Adaptation with an Interpretable Poker AI

Toshiki Ohhata[1,a)]　Yoshimasa Tsuruoka[1,b)]

Abstract: Solvers that compute Game Theory Optimal (GTO) equilibrium strategies are widely used for training in No-Limit Texas Hold'em (NLH), a leading variant of poker. Their outputs, however, are tables of how often each action should be taken with each hand: they tell players what to do without explaining why. Moreover, in a two-player zero-sum game, GTO is a static strategy designed to remain unexploitable regardless of the opponent, which is fundamentally different from actively exploiting the habitual mistakes (leaks) of the opponent at hand. Building on **advance0**, a manually designed rule-based AI that systematizes expert reasoning and structural GTO criteria as explicit rules, we construct (1) weak AIs that reproduce typical leaks along two axes—the range of hands played (loose/tight) and the level of aggression (aggressive/passive)—and (2) an exploitative bot that switches countermeasures according to an externally supplied opponent-type label. In our evaluation, advance0 exhibits agreement with GTO, and the exploitative bot significantly outperforms **base**, the opponent-agnostic baseline strategy, against three of the four weak-AI types. We further find that effective countermeasures differ across opponent types. These results demonstrate that interpretable rules can construct not only a GTO approximation but also opponent-adaptive exploitative strategies in a form that humans can trace.

Keywords: Poker AI, Imperfect Information Games, Interpretable AI, Game Theory Optimal (GTO), Opponent Modeling, Exploitative Strategy

[1] 東京大学大学院情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo
[a)] ohhata@logos.t.u-tokyo.ac.jp
[b)] tsuruoka@logos.t.u-tokyo.ac.jp

1. はじめに

チェス・将棋・囲碁などの完全情報ゲームでは，AI はすでに人間の最高水準を上回る性能に達している [12], [13].

© 2026 Information Processing Society of Japan

近年は不完全情報ゲームへ焦点が移り，2 人および多人数のノーリミットテキサスホールデム（NLH; No-Limit Texas Hold'em）において Libratus [3]，DeepStack [9]，Pluribus [4] がプロを上回る性能を達成した．これらは反実仮想後悔最小化（CFR; Counterfactual Regret Minimization）[15] に基づく自己対戦や再探索を主要な構成要素として，強い戦略を構築する．

現代のポーカーの学習環境では，GTO ソルバや学習サービスが広く普及している．しかしその出力は「各ハンド × 各アクションの頻度表」であり，たとえば「A♡K♠：ベット 67％／チェック 33％」のように，**何を**どの頻度で行うかは分かっても**なぜ**そうなるのかは読み取れない．戦略表そのものに加えて，最適手に至る**思考過程**にも大きな教育的価値があると我々は考える．

さらに，2 人対戦のゼロサムゲームにおける GTO 戦略は，「相手によらず長期的に搾取されない」ことを目的とした静的な均衡戦略であり，現実の対戦相手が持つ悪癖（リーク）を積極的に突く戦略とは本質的に異なる．実戦の相手はしばしば偏った戦略を取るため，学習者には「相手の型をどう見抜き，どう突くか」の理解も重要である．

本稿では，判断で参照した特徴量・閾値と 3 段階の推論系列を列挙でき，同じ入力から人間が最終行動まで追跡できることを**解釈可能**と定義する．理解度や記憶容易性そのものはユーザ実験を行っていないため評価対象としない．この定義の下で，著者らが手作業で構築したルールベース AI「**advance0**」を土台に，(1) 典型的リークを再現する 4 種類の**弱い** *AI*，および (2) 相手型ラベルに応じて対策を切り替える**エクスプロイト** *Bot* を新たに構築する．全ての判断は明示的なルールで構成されるため，各意思決定に対して自然言語による解説を生成できる点が本手法の特徴である．

以上を踏まえ，本稿は次の 2 つの問い（RQ; Research Question）に答えることを目指す．**RQ1**：人間が追跡できる明示的なルールだけで，GTO ソルバの行動選択をどこまで再現できるか．**RQ2**：同じルール体系の上で相手の型ごとの弱点を突く戦略を構成し，型を仮定しない基準戦略を上回る利得を得られるか．RQ1 に実験 1（§5），RQ2 に実験 2（§6）で答える．

本研究の貢献は次の 3 点である．(1) 解釈可能なルールベース AI のポストフロップ判断と GTO ソルバ出力を 6-max NLH の 3,478 ノード・132,983 ハンドという大規模で照合し，双方が共通に使う行動カテゴリ内で頻度配分を正規化した一致度 87.7％（正規化しない共通一致は 73.6％）を定量的に示したこと（*GTO* **模倣の精度の実証**）．(2) 同じルール体系の上に，典型的リークを誇張した 4 種類の弱い AI と，既知の相手型に応じて少数パラメータを方向付きで調整するエクスプロイト Bot を構築し，弱い AI に対し最大 +27.9 bb/100（100 ハンドあたりのビッグブラインド数）の利得を得て，4 アーキタイプのうち 3 型で基準戦略を有意に上回ったこと（**均衡から搾取への連続的拡張**）．(3) 基準戦略（base）と搾取戦略（exploit）の差分を自然言語で提示し，「相手の型に応じてなぜ・どう逸脱するか」を学習者が追跡できる形で示したこと（**解釈可能な搾取解説**）．

2. 背景

2.1 NLH と GTO 戦略

NLH では各プレイヤーに 2 枚の手札（ハンド）が配られ，プリフロップ・フロップ・ターン・リバーの 4 ストリートを通じて最大 5 枚の共有カード（ボード）を加えた役の強さを競う．各手番の行動は，賭けずに手番を渡す**チェック**，自分から賭ける**ベット**，相手と同額を出す**コール**，それを上回る額へ引き上げる**レイズ**，降りる**フォールド**，持ち点の全額を賭ける**オールイン**である．賭けられた金額の総和を**ポット**，降りずに残った者が手札を見せ合って勝敗を決めることを**ショーダウン**と呼ぶ．手番の順序を表す座席（ポジション）は SB（スモールブラインド），BB（ビッグブラインド），UTG（アンダーザガン），HJ（ハイジャック），CO（カットオフ），BTN（ボタン）で，後ろほど情報面で有利である．プレイヤーが取りうるハンドの集合を**レンジ**，ゲーム木上の意思決定点を**ノード**と呼ぶ．成績は 100 ハンドあたりの平均収支を強制賭け額の大きい方（ビッグブラインド，bb）1 枚で表した *bb/100* で示し，百分率どうしの差には *pp*（percentage point）を用いる．2 人対戦の NLH はゼロサムゲームであり，GTO 戦略はゲームのナッシュ均衡に対応し，CFR により近似計算される．この均衡戦略は相手の戦略によらず長期的に搾取されないが，相手が均衡から外れていても，それを**最大限に**突くわけではない．

2.2 エクスプロイトと最大搾取

相手の戦略 σ_{-i} が固定で既知のとき，それに対する利得を最大化する戦略をベストレスポンスと呼び [7]，本稿では**最大搾取戦略**（MES; Maximally Exploitative Strategy）と呼ぶ．GTO がリスク回避的な均衡であるのに対し，MES は相手の弱点を最大化的に突く．実用上は，相手の型を有限個の典型へ分類して各型に応じた対策を適用する近似が有効であり，本研究もこの方針を採る．たとえば相手が過剰にブラフする型であれば，GTO が守る最小防御頻度（MDF; Minimum Defense Frequency）を意図的に上回ってコールを増やすことで，均衡では取りこぼすブラフ部分の期待値まで回収できる．逆に相手が強い手でしか打たない型であれば，MDF を下回って降りることで無駄なコールを避けられる．このように MES 的な調整は「相手の閾値のずれと逆方向に自分の閾値をずらす」という原理に帰着でき，少数のパラメータ操作として実装しやすい．本研

© 2026 Information Processing Society of Japan

究のエクスプロイト Bot は，この原理を advance0 の**ポットオッズ**マージン，MDF，バリュー頻度の方向付き調整として具体化する．ポットオッズはコールに必要な額をコール後のポット全体で割った値であり，これを上回る勝率が見込めればコールが有利になる．

2.3 advance0 の概要

advance0 は著者らが本研究に先立って実装したエキスパートシステムであり，熟練者の段階的思考，ポーカーの定石，バリューとブラフの比（VB 比）や MDF 等の GTO の構造的基準を手作業でルールへ落とし込んだものである．ルールの自動抽出手法ではなく作成者の知識に依存するため，作成コストと作成者間再現性は本研究の限界である．判断は以下の 3 段階で行う．

Step 1. ハンド評価　相手レンジに対するエクイティ（EQ; equity．手札が最終的に勝つ確率，%），ドローポテンシャル（あと 1 枚で強い役が完成する見込み），ブロッカー効果（自分のカードが相手の強い役の存在を阻害する効果）を算出する．

Step 2. バンド分類　ハンドを VALUE（強い；バリュー可），SEMI（ドロー；将来性あり），MARGINAL（中程度；チェック／コール主体），BLUFF（弱いがブロッカー良好）の 4 バンドへ分類する．

Step 3. 頻度決定　ポットタイプ（プリフロップで何回レイズが入ったか），ストリート，ベットサイズに応じて各アクションの頻度を決定する．

頻度決定では，ベットサイズ r（ポット比）に対するバリュー対ブラフ比 VB 比 $= 1 : \frac{r}{1+r}$ や，最小防御頻度 MDF $= \frac{1}{1+r}$ を基準に用いる．ベットサイズは表 1 のように離散化する．各ステップが明示的なルールであるため，参照特徴 → 中間分類 → 行動という思考フロー全体を同じ順序で自然言語として提示できる（図 1）．例えば K のトップペアなら，EQ・ドロー・ブロッカーを算出し，VALUE に分類し，相手レンジと VB 比から small bet の頻度を決めて，「K のトップペアで VALUE．相手のセカンドペアやドローから得るため small bet」と出力する．



図 1　advance0 の 3 段階パイプライン

3. 関連研究

Rudin [11] は高リスク意思決定における本質的な解釈可能性の重要性を論じる．CFR や深層強化学習を用いた高性能ポーカー AI [2], [9], [14] は高い対戦性能を示す一方，戦略が数値モデルとして表現され，人間が判断根拠を直接追跡することは難しい．これに対し，Ganzfried ら [6] は人間が実行可能なルールを導出し，Bertsimas ら [1] は決定木と CFR により，解釈可能で世界水準の対戦性能を持つ Bot を構築した．PioSOLVER [10] 等の GTO ソルバは解いた戦略（行動頻度）や期待値を表示するが，本研究のような明示的な判断手順を直接与えるものではない．また様々な棋力の人間の再現という観点では，人間棋譜で学習したチェス AI の Maia [8] が既存エンジンより高精度に人間の着手を予測した．MES は強力だが強い相手に逆に搾取される危険があるため，その危険を制限する安全な搾取（safe opponent exploitation）[5] も研究されてきた．本研究は相手を少数の解釈可能なアーキタイプに分類し，リークの所在と対策を明示的に結び付けることで，弱者の再現から搾取の根拠までを解釈可能なルールで一貫して提示する点で異なる．

表 1　ベットサイズのカテゴリ（ポット比）

カテゴリ	ポット比
small / medium / large	0–40 / 40–90 / 90–140 %
overbet / overbet+ / all-in	140–200 / 200 %以上 / 全額

4. 提案手法

以下，相手の型を仮定しない advance0 の既定設定を**base（基準戦略）**と呼ぶ．base は実験 1（§5）で GTO との一致度を測った戦略そのものであり，これに型別の逸脱を加えた戦略を **exploit（搾取戦略）**と呼ぶ．以後の比較はすべて base を基準とする．

4.1 弱い AI：4 アーキタイプ

ポーカーで頻出する悪癖を，参加レンジの広狭（ルース／タイト）と攻撃性の高低（アグレッシブ／パッシブ）の 2 軸で 4 象限に分類し，それぞれを LAG・LP・TAG・TP（L はルース，T はタイト，A はアグレッシブ，P はパッシブ）と呼んで，advance0 のパラメータ上書き（overrides）として実装した（表 2）．LAG は参加レンジを広げ，ブラフ・ベット転換・根拠の薄いレイズを大幅に引き上げる．LP は広く参加するがブラフでの再レイズを封印し，ポストフロップのブラフを停止してコール中心に振る舞う．TAG は参加を 1 段狭めつつブラフでの再レイズとセミブラフ（完成前の手での攻撃）を増やし，TP は参加を狭め，強い手のみ打ち，受けは過度に降りる．これらは「弱くした AI」ではなく，特定のリークを誇張した**人間らしい**相手の再現を意図する．各型は mild ／ severe のバリアントを持ち，リーク強度を調整できる．

© 2026 Information Processing Society of Japan

表 2　弱い AI の 4 アーキタイプ

型	主なリーク
LAG	広く参加・過剰なブラフ／ベット
LP	広く参加・コール過多・ブラフ皆無
TAG	狭く参加・過剰な攻撃
TP	狭く参加・強い手のみ・過度な降り

4.2　型ラベル既知のエクスプロイト Bot

エクスプロイト Bot は外部から与えられた相手のアーキタイプに応じて advance0 のパラメータを切り替え，各型のリークを突く（表 3）．対策の方向性は MES の考え方に基づく．vs LAG はブラフが多いためポットオッズマージンを緩めコールを厚くし，MDF を上げてブラフを捕える．vs LP は降りないためバリュー頻度を上げ薄いバリューまで取り切る．vs TAG はポットオッズと MDF を緩めて受けを広げる．vs TP は強い手でしか攻めてこないため相手のベット・レイズには多めに降り，降りずに付いてくる相手へは薄いバリューまでベットする．各対策は十数件規模の ablation（後述の対戦）で選定した．TP 向けは，誤判定の分析（§6.5）で LP 向けの対策の方が TP に効くと分かったため，評価に用いない seed 10600–10799 で候補を選び直した．本研究では対策自体の効果を分離するため正しい型ラベルを既知とし，型分類器は Bot へ組み込んでいない．したがって結果は型別適応の上限評価であり，未知・中間型や動的に戦略を変える相手の性能を保証しない（型ラベルが現実に得られるか，誤判定がどれだけ高くつくかは §6.5 で別途測る）．重要なのは，これらの対策がいずれも advance0 の既存パラメータの**方向付きの上書き**として表現される点である．新たなゲーム木探索や学習を必要とせず，「どのしきい値を，どちらの向きに，どれだけ動かすか」という少数の操作で搾取が構成されるため，exploit は常に base からの差分として解釈でき，リーク強度に応じた段階調整も容易になる．

表 3　エクスプロイト Bot の対策方針

相手	対策	主な調整
LAG	最後まで受ける	ポットオッズ緩和・MDF ↑
LP	バリューを厚く	バリュー頻度↑・薄バリュー
TAG	緩く受ける	ポットオッズ・MDF 緩和
TP	攻めには降り薄く取る	薄バリュー↑・降り↑

4.3　自然言語による差分解説

エクスプロイト時には，同一局面で「base であればどう打つか」を影の戦略として並行計算し，base の手と exploit の手の差分種別を分類する．これにより**なぜ通常と異なる判断をするのか**を自然言語で提示でき，学習者は base と exploit を同じ局面で並べて，搾取がどのリークを前提にどう逸脱するのかを追える．たとえば，あるボードでマージナルなハンドを持つ局面では，次のような差分解説が得られる．

> **base**：チェックが基本．ショーダウンまで持ち込める価値を残しつつ受けに回る．
> **exploit（vs LAG）**：相手は高頻度でブラフしてくるので，ここはチェックしてベットを誘い，リバーまでコールで捕える．

このように，base と exploit の双方とその理由が同時に提示されるため，学習者は「相手がこういう型だからこう変える」という思考の写像を獲得できる．

5.　実験 1：GTO との一致度（RQ1）

5.1　設定と指標

6-max NLH・100BB において，advance0 のポストフロップ判断を GTO ソルバの出力と比較した（3,478 ノード・132,983 ハンド，フロップ・ターン・リバー）．ベット／チェックと受け（フォールド／コール／レイズ）を対象とし，ベットサイズは表 1 のカテゴリへ離散化する．一致度には，2 つの頻度分布の重なり（$\sum_c \min(a_c, g_c)$）に基づく**頻度一致**，「5 %以上使うか」の二値判定一致率である**閾値一致**，および EQ 推定の平均絶対誤差（MAE; Mean Absolute Error）と決定係数 R^2 を用いた．頻度一致は 3 段階で見る．*all* は全カテゴリ，*com* は両者が共通に使うカテゴリのみ，*com_norm* はさらに共通部分内で頻度を正規化した値である．all と com の差は「使うサイズの集合が一致するか」を，com と com_norm の差は「その集合内で配分が一致するか」を表すため，乖離がサイズ選択と頻度配分のどちらに由来するのかを切り分けられる．EQ 推定は，advance0 が Step 1 で算出する相手レンジに対する EQ と，GTO ソルバが同一局面で用いる EQ を，双方のレンジを揃えて全ハンドにわたり比較した．対象はプリフロップのアクション系列で定義される代表的ポストフロップ局面の 1001–5000 番である．

5.2　全体・ストリート別の結果

全体の一致度を表 4 に，ストリート別を表 5 に示す．双方が共通に使うカテゴリ内で頻度配分を正規化した一致度は 87.7 %，正規化しない共通一致は 73.6 %，閾値一致は 89.1 %であり，EQ 推定は $R^2 = 0.86$（MAE 7.11 pp）であった．これは advance0 が，共通して使用するカテゴリ内では配分も GTO に近いことを示す．一方，全カテゴリを含む一致は 61.5 %であり，極端サイズの採否まで含めた完全な GTO 再現を意味しない．全体のベット頻度は advance0 44.0 %対 GTO 42.3 %（差 +1.7 pp）でわずかに攻撃的である．ストリート別ではフロップで EQ 推定が最も正確（MAE 5.00 pp），リバーで最難（10.74 pp）である一方，共通正規化頻度はリバーで最高（93.2 %）であり，リ

© 2026 Information Processing Society of Japan

バーでは「どのサイズを使うか」は揃うが「どのハンドで打つか」の推定が難しい傾向が読み取れる．

MAE 7.11 pp の水準を示すため単純な推定量と比べると，局面は分かるが手札が分からない推定（そのノードの GTO EQ の平均値を全ハンドへ与える）は MAE 22.81 pp，$R^2 = 0.08$ であり，advance0 は 3.2 倍正確である．EQ 推定は局面の平均的な強さではなく手札ごとの強さを捉えているといえる．また，GTO 側の最大行動頻度で局面を層別すると，ほぼ単一行動に寄る自明な局面（最大頻度 90％以上，77,066 ハンド）で com_norm 93.6％，行動が割れる判断の要る局面（同 70％未満，33,248 ハンド）でも 79.0％であり，一致は自明な局面だけで稼いでいるわけではない．

表 4 GTO との全体一致度

指標	all	com	com_norm
Freq agreement	61.5％	73.6％	87.7％
Threshold agreement	64.6％	89.1％	—
EQ MAE / R^2	7.11 pp / 0.86		

表 5 ストリート別の一致度

Street	ハンド数	f(all)	f(com_norm)	EQ MAE
Flop	54,404	66.9％	88.9％	5.00 pp
Turn	44,661	60.9％	82.0％	6.92 pp
River	33,918	53.7％	93.2％	10.74 pp

5.3 バンド・サイズ・受けの内訳

バンド別ベット頻度（表 6）では，VALUE，SEMI で advance0 がやや高頻度に打つ一方，MARGINAL，BLUFF は GTO にほぼ一致し，バンド構造が再現できている．ベットサイズ分布（表 7）では advance0 は medium を多用し small がやや少なく，overbet やオールインの利用が GTO より少ない（ルール化に伴うサイズ集合の簡略化を反映）．各サイズの VB 比も GTO と近い．

表 6 バンド別ベット頻度（advance0 vs GTO）

バンド	advance0	GTO	差
VALUE	62.4％	58.7％	+3.7
SEMI	51.8％	46.2％	+5.6
MARGINAL	31.9％	31.0％	+0.9
BLUFF	37.1％	38.8％	−1.7

受け（923 ノード）では，フォールド頻度が GTO とほぼ一致し（33.5％対 33.0％），コールがやや低くレイズがやや高いことから，受けの総量（MDF）は保ったまま内訳だけがやや攻撃寄りになっていると解釈できる．ポットに残る人数別では，人数が増えるほど全カテゴリ一致は下がるが（2 人 63.5％から 4 人 57.9％），共通正規化頻度は 85.8–89.5％と高く保たれる．

表 7 ベットサイズ分布（advance0 vs GTO，単位は%）．ob は overbet，ob+は overbet+，ai は all-in（表 1 のカテゴリ）

	small	medium	large	ob	ob+	ai
advance0	30.3	37.0	7.7	0.4	0.1	1.3
GTO	35.8	32.7	5.0	2.7	1.3	6.6

5.4 GTO 模倣の解釈：なぜ一致し，どこで乖離するか

一致が得られた要因は，advance0 が GTO と同じ**構造的基準**を明示的に持つことにあると考えられる．VB 比と MDF はいずれも GTO 戦略の均衡条件に対応する量であり，これらをルールの中核に据えることで，個々のハンドの頻度を網羅的に学習せずとも均衡戦略の骨格が再現される．バンド別頻度で MARGINAL，BLUFF が GTO にほぼ一致するのはこの効果であり，VALUE，SEMI でわずかに高頻度なのは，ルールが薄いバリューを取りこぼさないようやや攻撃側に倒していることを反映する．

乖離は主に 2 つの軸に現れる．第一は**サイズ集合の簡略化**であり，ルール数を抑え解説を平易に保つための意図的な単純化である．頻度的に支配的な small〜medium を押さえれば均衡戦略の大部分は再現でき，共通正規化頻度が高く保たれることから，この簡略化による損失は限定的だと考えられる．第二は**ストリート進行に伴う EQ 推定の困難化**で，リバーではレンジが細く分岐するため MAE が 10.74 pp に増大する．ただし「どのサイズを使うか」はリバーで最も一致する（com_norm 93.2％）ため，乖離は**ハンド単位の EQ 値**の推定に集中しており，リバー特化の EQ 補正が改善の主軸になると示唆される．なお一致率は，最大搾取される側に立ったときの損失量（exploitability）を直接測る指標ではなく，advance0 が均衡に近い強さを持つことまでは示さない．

6. 実験 2：弱い AI とエクスプロイト対戦（RQ2）

6.1 設定

弱い AI（§4）とエクスプロイト Bot をヘッズアップ（HU）で対戦させ，100 ハンドあたりの利得（bb/100）を測定した．HU を採用するのは，多人数卓では複数の相手型が同時に存在し利得を相手別に切り分けにくいためである．分散を抑えるため，同一のハンド列でポジション（BTN／BB）のみを入れ替えて 2 回対戦させ（seat-swap），カードの引きによる分散を相殺して戦略差に由来する利得を抽出する．各型について 10,000 ハンドで対策版十数件を比較して候補を絞り，最終候補を同一 100,000 ハンド ×2 席，計 200,000 ハンド／条件で再評価した．利得には標準誤差（SE）を併記する．改善量（exploit−base）は，同じ seed

© 2026 Information Processing Society of Japan

のハンド列を base と exploit の両方に配り，seed ごとの差 d_i の平均として求めた（SE は d_i の標準偏差 $/\sqrt{n}$）．同じ配札では打ち方が一致したハンドの損益が等しく差が 0 になるため，差の分散は打ち方が分かれたハンドだけから生じる．TP 相手の 10,000 ハンドでは損益が一致したハンドが 93.5 %であり，1 ハンドあたりの差の標準偏差は 4.3 bb と，各々の損益の標準偏差（12.5 bb）より小さく，差の分散は独立に対戦させた場合の約 6 %であった．

6.2 結果

各型に対するエクスプロイト Bot の利得を表 8 に示す．改善量（exploit−base）は 4 型すべてで点推定が正で，LP に対する絶対利得 +27.90 bb/100 が最大であった．95 %信頼区間（±1.96 SE）が 0 を含まないのは LAG・LP・TP で，TAG は含む．そこで追加 seed 数を事前に 1,200 と定め，構成の選択に用いていない seed 9400–10599 で同条件の対応比較を行った．計 1,600 seed で TP は +4.93 ± 0.60 bb/100 と有意であり（追加分のみでも +5.15 ± 0.69），TAG は +0.66 ± 0.43 bb/100 で信頼区間が 0 を含んだままであった．すなわち exploit が base を有意に上回るのは LAG・LP・TP の 3 型であり，TAG では有意差を示せていない．base は 4 型すべてに対して正の利得を得ており，「基準戦略はどの弱い AI にも負けない」という要件を満たす．これは弱い AI の強度を校正した結果であり，特に LAG は過去の過剰に強い版（base が負けていた）から参加率を下げた版へ差し替えている．同一カード列 400 点での累積利得の推移を図 2 に示す．exploit の優位は特定の区間に偏らず，全区間を通じて安定して現れている．

表 8 弱い AI に対する対応比較（bb/100）．1 条件あたり同一 400 seed×2 席，計 200,000 ハンド．* は seed を追加した計 1,600 seed（同一条件）

	LAG	LP	TAG	TP
base	+5.88	+21.49	+6.42	+12.02
exploit	+11.80	+27.90	+7.77	+16.27
exploit−base	+5.92	+6.41	+1.36	+4.25
対応差 SE	±1.32	±1.53	±0.93	±1.22
exploit−base*	—	—	+0.66	+4.93
対応差 SE*	—	—	±0.43	±0.60

6.3 型別の搾取メカニズム

利得の大小は各型のリーク構造から解釈できる．LP で改善量と絶対利得が最大なのは，広くコールするリークに対してバリューベットとレイズを増やし薄いバリューまで取り切る調整が直接作用したためである．LAG と TAG はともに攻撃が過剰であり，こちらが受けを広げることでブラフ部分を回収できる．TP は強い手でしか攻めない一方で降りずに付いてくるため，攻めには多めに降り，付いてくる手へ薄いバリューを打つ調整が有効であった．このように本枠組みは「型 → リークの所在 → 対応する閾値調整」という解釈可能な対応付けを与え，搾取が効く／効かない理由まで言語化できる．

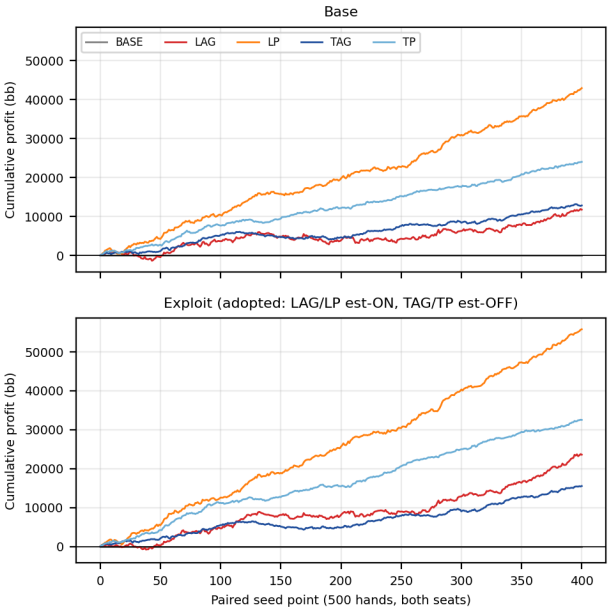


図 2 同一カード列 400 点（各 500 ハンド）での累積利得．上=base，下=exploit．全型で exploit（下）が base（上）を上回り，特に LAG/LP で顕著．seed 9000–9399 の対応比較

6.4 6max 卓での検証：デュプリケート測定

HU の結果が 6 人卓でも成り立つかを，配札運を消した測定法で検証した．6max・100BB で，1 つの配札を 6 回プレイし，セット内ではボタンを固定したまま座席の割り当てだけを 1 つずつ回す（**デュプリケート**）．6 回で各プレイヤーが同じ手札と全ポジションを 1 回ずつ経験するため，配札・ポジション運を相殺できる．卓は *1* **種類のアーキタイプのみ**を含む構成（base 4 体＋対象 2 体）とし，席順はセットごとにシャッフルした．4 型それぞれ 12,000 セット，計 288,000 ハンドを測定した．

結果を表 9 に示す．参加率 VPIP はルース型（LAG 49.6 %，LP 33.8 %）がタイト型（TAG 22.1 %，TP 19.8 %）より高く，攻撃性 AggF（フロップでベットまたはレイズを選ぶ率）はアグレッシブ型（LAG 43.0 %，TAG 36.2 %）がパッシブ型（LP 27.0 %，TP 26.3 %）より高い．2 軸が定義通りに分離しており，4 アーキタイプが狙ったリーク構造を持つことが裏づけられる．利得は base が 4 型すべてに対して正（+9.60, +6.60, +0.96, +0.22 bb/100）で，「基準戦略はどの弱い AI にも負けない」という要件が 6max でも成立する．ただし TAG・TP に対する優位は信頼区間が 0 を含み，有意ではない．

ショーダウンに至った手の勝率（hW$SD）は，アグレッシブ型ほど低く（LAG 44.2 %，TAG 48.0 %），パッシブ型

© 2026 Information Processing Society of Japan

ほど高い（LP 48.8％，TP 51.5％）．ブラフを含む広いレンジでショーダウンする型ほど勝ち切れないという解釈と整合する．

表 9 6max デュプリケート測定（4 型 ×12,000 セット，計 288,000 ハンド）．VPIP は参加率，AggF はフロップでベットまたはレイズを選ぶ率，hW$SD は 1 対 1 でショーダウンに至った手の勝率（引き分けを除く）で単位は%．利得は base 視点（正なら base の勝ち，± は 95％信頼区間）で，base の統計は 4 卓合算

型	VPIP	AggF	hW$SD	base の利得
base	21.5	31.4	51.5	—
LAG	49.6	43.0	44.2	+9.60 ± 3.57
LP	33.8	27.0	48.8	+6.60 ± 2.65
TAG	22.1	36.2	48.0	+0.96 ± 1.28
TP	19.8	26.3	51.5	+0.22 ± 1.07

6.5 型ラベルの実現可能性と誤判定コスト

実験 2 は正解型を既知とするため，ラベルが現実に得られるのか，外したときにどれだけ損をするのかを別途測った．外側から観測できる統計のみで型を 4 クラスに分類すると（チャンスレート 25.0％，系列単位の 5 分割交差検証），参加率と攻撃性の 2 軸ルールで 500 ハンド 86.7％・2,000 ハンド 95.0％，16 項目・深さ 3 の決定木で 500 ハンド 95.8％・2,000 ハンド 98.3％であり（図 3），型ラベルは数百ハンドの観測から得られる見込みがある．2 軸ルールの 500 ハンド時点の誤り 16 件（120 系列中）のうち 12 件は TAG を TP とするもので，200 ハンド以上で LAG を取り違えた例はなく，後述する最も高くつく誤りは起きていない．

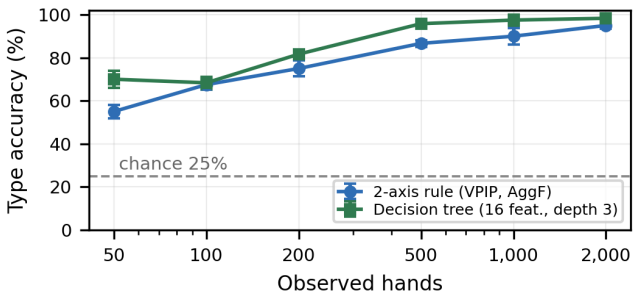


図 3 観測ハンド数と型推定の正解率（120 系列，5 分割交差検証，誤差棒は標準誤差）

また，誤った型を信じさせた 12 通りを同一配札 400 seed で測ると，base より有意に損をするのは 2 通りだけで，いずれも真の相手が LAG の場合であった（TP と誤認 −5.43，LP と誤認 −3.59 bb/100）．攻撃的な相手を受動的な型と取り違える誤りが最も高くつく．一方，LP 向けと TP 向けの対策はどちらも，薄いバリューを厚くする点で共通しており，真の相手が LP でも TP でも互いに有意に得をした．

表 10 型を誤って信じたときの改善量（exploit−base, bb/100）．同一 400 seed×2 席．対角が正解．± は 95％信頼区間の半幅（1.96× 対応差 SE）．太字は 0 から有意に離れるセル．LAG・LP 向けは相手レンジ推定あり，TAG・TP 向けはなし（各型で採用した構成）

信じた型	真の相手			
	LAG	LP	TAG	TP
LAG	**+5.92**±2.58	−0.59±2.22	+0.23±2.00	−1.48±1.83
LP	**−3.59**±3.27	**+6.41**±2.99	−0.09±2.59	**+4.51**±2.19
TAG	+0.93±2.16	−0.35±1.87	+1.36±1.82	−0.75±1.42
TP	**−5.43**±3.41	**+5.65**±3.02	+1.60±2.53	**+4.25**±2.39

12 通りの誤判定のうち，base を有意に下回るのは 2 通りのみであり，搾取戦略は型を外しても大きくは崩れない．また各列で対角は最大か，最大との差が信頼区間内であった（TAG 列 −0.24，TP 列 −0.26）．

7. 考察

GTO 模倣と搾取の連続性．エクスプロイト Bot は新たな戦略を一から学習するのではなく，base が用いるポットオッズマージン・MDF・バリュー頻度を相手の型に応じてずらすだけで構成される．したがって exploit は常に「base からの差分」として表現でき，均衡からどれだけ・どの方向に逸脱しているかが常に明示される．これはブラックボックスな相手モデリングでは得られない性質であり，学習者が「安全な均衡」と「相手依存の逸脱」を地続きに理解する助けになる．

解釈可能性のコストと利点．ルール化に伴うサイズ集合の簡略化（overbet ／オールインの不足）やリバー EQ の誤差は，解釈可能性のために支払うコストである．それでも共通に使うカテゴリ内では配分を 87.7％まで保てており，全判断を言語化できる利点はこのコストに見合うと考える．数値的な再現度は CFR 系 AI に及ばない一方，本手法は**なぜその手なのか**を一貫して提示できる．

課題と限界．(1) リバーの EQ 推定精度（MAE 10.74 pp）とサイズ分布の偏り（medium の多用，overbet ／オールインの不足）が残る．少数サイズに集約したルールでは極端サイズを要する均衡バランスを完全には再現できないため，リバー専用の EQ 補正と，特定ボードでの overbet 許可ルールの導入が考えられる．(2) TAG への改善量は 1,600 seed でも信頼区間が 0 を含み，候補探索の精緻化が必要である．(3) 弱い AI はリークを誇張した近似であり，実プレイ履歴と各アーキタイプの統計を突き合わせて型の妥当性を検証することが望ましい．(4) 型ラベルの推定は外側の統計を用いたオフライン分類として確かめたにとどまり（§6.5），Bot へ組み込んだ状態での精度，中間型，動的に戦略を変える相手は評価していない．(5) 行動頻度の一致は戦略強度を保証しないため，近似ベストレスポンスによ

© 2026 Information Processing Society of Japan

る exploitability 評価が必要である．

教育・学習支援への応用． base と exploit の差分を自然言語で示すことで，学習者は均衡と搾取の使い分けを追跡できる．またプレイ履歴を型のパラメータ空間へ射影すれば，学習者自身のリーク診断への拡張も可能である．

8. まとめと今後の展望

解釈可能なルールベース AI advance0 を土台に，典型的リークを再現する 4 種類の弱い AI と，既知の相手型に応じて対策を切り替えるエクスプロイト Bot を構築した．

RQ1 への回答． 明示的なルールだけでも，GTO ソルバと共通に使う行動カテゴリ内では頻度配分を 87.7％（正規化しない共通一致 73.6％）まで再現でき，バンド構造・VB 比・MDF に基づく受けも GTO に近い．一致は自明な局面に偏っておらず，行動が割れる局面でも 79.0％である．一方，全カテゴリを含む一致は 61.5％であり，overbet やオールインの採否までは再現できない．すなわち再現できるのは均衡戦略の骨格までであり，一致率は exploitability を意味しない．

RQ2 への回答． 同じルール体系の上で，少数パラメータの方向付き上書きだけで搾取戦略を構成でき，弱い AI に対して最大 +27.9 bb/100 の利得を得た．base との差が有意なのは LAG・LP・TP の 3 型（+5.92, +6.41, +4.93 bb/100）で，TAG は点推定では正だが有意でない（$+0.66 \pm 0.43$）．base は 4 型すべてに正の利得を得ており，基準戦略がどの弱い AI にも負けないことも確認した．有効な対策は型ごとに異なる．6 人卓でも配札運を相殺するデュプリケート測定により，2 軸のリーク構造が定義通りに分離していることを裏づけた．ただし本実験は正解型を既知とする上限評価である．

以上より，GTO 模倣（実験 1）とエクスプロイト（実験 2）が同一のルール語彙の上で連続的に構成でき，搾取が効く理由・効かない理由までを型とリークの対応として言語化できることを確認した．数値的には精緻でも判断根拠を直接追跡しにくい CFR 系 AI とは異なり，均衡から搾取への移行を学習者が追跡可能な形で扱える点に本手法の意義がある．

今後は，(1) 型分類器を組み込んだ状態での評価と，中間型・動的相手に対する頑健性の検証，(2) 近似ベストレスポンスによる advance0 の exploitability 評価，(3) TAG 向けエクスプロイトの追加探索，(4) リバー特化の EQ 補正，(5) プレイ履歴を型のパラメータ空間へ射影する弱点自動検出に基づく学習支援に取り組み，解釈可能な AI による「均衡と搾取の統一的な学習支援」の実現を目指す．

参考文献

[1] Bertsimas, D. and Paskov, A.: World-class interpretable poker, *Machine Learning*, Vol. 111, No. 8, pp. 3063–3083 (2022).

[2] Brown, N., Bakhtin, A., Lerer, A. and Gong, Q.: Combining Deep Reinforcement Learning and Search for Imperfect-Information Games, *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 33, pp. 17057–17069 (2020).

[3] Brown, N. and Sandholm, T.: Superhuman AI for heads-up no-limit poker: Libratus beats top professionals, *Science*, Vol. 359, No. 6374, pp. 418–424 (2018).

[4] Brown, N. and Sandholm, T.: Superhuman AI for multiplayer poker, *Science*, Vol. 365, No. 6456, pp. 885–890 (2019).

[5] Ganzfried, S. and Sandholm, T.: Safe Opponent Exploitation, *ACM Transactions on Economics and Computation (TEAC)*, Vol. 3, No. 2 (2015).

[6] Ganzfried, S. and Yusuf, F.: Computing Human-Understandable Strategies: Deducing Fundamental Rules of Poker Strategy, *Games*, Vol. 8, No. 4, p. 49 (2017).

[7] Johanson, M., Waugh, K., Bowling, M. and Zinkevich, M.: Accelerating Best Response Calculation in Large Extensive Games, *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 258–265 (2011).

[8] McIlroy-Young, R., Sen, S., Kleinberg, J. and Anderson, A.: Aligning Superhuman AI with Human Behavior: Chess as a Model System, *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1677–1687 (2020).

[9] Moravčík, M., Schmid, M., Burch, N., Lisý, V., Morrill, D., Bard, N. et al.: DeepStack: Expert-level artificial intelligence in heads-up no-limit poker, *Science*, Vol. 356, No. 6337, pp. 508–513 (2017).

[10] PioSOLVER: GTO Poker Solver, https://piosolver.com/. Accessed: October 2026.

[11] Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, *Nature Machine Intelligence*, Vol. 1, No. 5, pp. 206–215 (2019).

[12] Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G. et al.: Mastering the game of Go with deep neural networks and tree search, *Nature*, Vol. 529, pp. 484–489 (2016).

[13] Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A. et al.: A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play, *Science*, Vol. 362, No. 6419, pp. 1140–1144 (2018).

[14] Zhao, E., Yan, R., Li, J., Li, K. and Xing, J.: AlphaHoldem: High-Performance Artificial Intelligence for Heads-Up No-Limit Poker via End-to-End Reinforcement Learning, *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36, No. 4, pp. 4689–4697 (2022).

[15] Zinkevich, M., Johanson, M., Bowling, M. and Piccione, C.: Regret Minimization in Games with Incomplete Information, *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 20 (2007).

© 2026 Information Processing Society of Japan