アタック25における手詰まり盤面の戦術均衡解析

浜向　直[1,a)]　阿原　一志[2,b)]

概要：本研究では、クイズ番組『パネルクイズ アタック 25』において、すべてのプレイヤーがクイズへの解答を避けることが合理的となる「手詰まり盤面」の存在を検討した。各プレイヤーが Pass (P) または Answer (A) を同時に選択し、Answer を選択したプレイヤーの中から確率的に手番者が決定されるゲームとしてモデル化した。盤面評価には、多人数確率手番ゲームに対応した UCT エージェントを用い、各盤面について 100 回の評価を行った推定勝率の平均値を用いた。単色十字型盤面を対象として各プレイヤーの最適反応を調べた結果、他の 3 人が Pass を選択しているとき、すべてのプレイヤーにとって Pass が Answer より高い推定勝率を与え、全員 Pass の戦略プロファイル (P, P, P, P) が狭義ナッシュ均衡となることを確認した。さらに、Pass/Answer の 16 通りの戦略プロファイルから推定利得ゲームを構成して解析した結果、全員 Pass が唯一の相関均衡となることを示した。Regret Matching による学習シミュレーションにおいても、経験分布は全員 Pass に集中した。以上により、アタック 25 に手詰まり盤面が存在することを示した。

キーワード：パネルクイズアタック 25, 多人数不確定完全情報ゲーム, UCT エージェント

Strategic Equilibrium Analysis of a Stalemate Board Position in Attack 25

Nao Hamamuki[1,a)]　Kazushi Ahara[2,b)]

Abstract: This study investigated whether a stalemate position exists in the four-player game Panel Quiz Attack 25, in which it is rational for all players to refrain from answering a quiz question. The game was modeled as a combination of a simultaneous-move game, in which each player chooses either Pass (P) or Answer (A), and a stochastic turn-taking game, in which one of the players who chose Answer is selected probabilistically to take the turn. We evaluated board positions using a UCT agent designed for multiplayer stochastic turn-taking games. We used the mean estimated winning probability from 100 evaluations of each position as its evaluation value.

For a monochromatic cross-shaped position, we examined each player's best response. The results showed that, when the other three players chose Pass, Pass yielded a higher expected winning probability than Answer for every player. Thus, the all-Pass strategy profile (P, P, P, P) constituted a strict Nash equilibrium. We further constructed an estimated payoff game from the 16 possible Pass/Answer strategy profiles and showed that the all-Pass profile was its unique correlated equilibrium. In a learning simulation using Regret Matching, the empirical distribution also became highly concentrated on the all-Pass profile. These results demonstrated a stalemate position in Attack 25.

1. はじめに

『パネルクイズ アタック 25』（以下、アタック 25）は、赤 (R)、緑 (G)、白 (W)、青 (B) の 4 人の出場者がクイズに解答し、正解することで得られるパネル獲得権を用いて盤上のパネルを取り合うクイズ番組である。番組終了時に最も多くのパネルを獲得した出場者が勝者となる。アタック 25 の基本ルールおよびゲーム構造については、阿原ら [1] を参照されたい。

野口ら [2] は、アタック 25 からクイズ要素を取り除き、4 人のプレイヤーに対してパネル獲得権（以下、手番）が確率的に与えられるゲームとして解析を行った。

[1] 北海道大学大学院理学研究院数学部門
Hokkaido University
[2] 明治大学総合数理学部
Meiji University
[a)] hnao@math.sci.hokudai.ac.jp
[b)] ahara@meiji.ac.jp

© 2026 Information Processing Society of Japan

同研究では、モンテカルロ法を用いて各盤面におけるプレイヤーの勝率を推定し、盤面によっては、プレイヤーが意図的にクイズへの参加を見送り、手番を取得しない方が自己の勝率を高く保てる場合があることを指摘している。一方、4 人のプレイヤー全員が自発的に手番の取得を避けるような盤面については、その存在が未解決の問題として残されていた。

本研究では、クイズに解答しないことを Pass(P)、クイズに解答して手番の取得を目指すことを Answer (A) と呼ぶ。各プレイヤーが Pass または Answer のいずれかを選択すると考え、他の 3 プレイヤーが Pass を選択しているときに、各プレイヤーにとって自らも Pass を選択することが最適反応となる盤面、すなわち戦略プロファイル (P, P, P, P) がナッシュ均衡となる盤面を、本論文では**手詰まり盤面**と呼ぶ。この定義は、Pass が他のプレイヤーの任意の行動に対して常に最適となることを要求するものではない。

手詰まり盤面の問題を扱うため、本研究ではアタック 25 を同時手番ゲームと確率手番ゲームの組合せとして捉える。クイズが出題された時点で、各プレイヤーは他のプレイヤーの選択を知らないまま Pass または Answer を選択するものとする。この段階は 4 人の同時手番ゲームとみなすことができる。その後、Answer を選択したプレイヤーの中から一定の確率で 1 人が手番を取得し、パネルを選択する。この段階を確率手番ゲームとして扱う。

本研究では、まず UCT を用いた多人数確率手番ゲーム用エージェントによって、典型的な単色十字型盤面の評価を行う。次に、盤面の対称性を用いて各プレイヤーの合法手を分類し、同一盤面を繰り返し評価して得られた推定勝率の平均値を用いて、各プレイヤーの最適反応を調べる。その結果から、全員 Pass の戦略プロファイルが狭義ナッシュ均衡となることを示す。さらに、Pass/Answer の 16 通りの戦略プロファイルから推定利得ゲームを構成し、相関均衡の構造を解析するとともに、Regret Matching による学習シミュレーションを行う。これらの解析により、アタック 25 に本論文の意味での手詰まり盤面が存在することを示す。

2. UCT エージェントによる盤面評価

2.1 本研究で用いる UCT エージェントについて

著者らが開発したアタック 25 ボードゲームシミュレータ [3] には、Monte Carlo Tree Search（MCTS）に基づく複数の探索エージェントが実装されている。本研究では、UCT（Upper Confidence bounds applied to Trees）を用いた D4E100P1maxn エージェントを盤面評価に使用する。

UCT では、各ノードにおいて、これまでのプレイアウトから得られた評価値と探索回数に基づいて UCT 値を計算し、その値に従って次に探索するノードを選択する [4]。これにより、評価値の高い有望な分岐に探索資源を重点的に配分する一方、探索回数の少ない分岐についても探索の機会が与えられる。選択された葉ノードから先については、ゲーム終了までパスを含まないランダムプレイによるプレイアウトを行い、その結果を探索木へ逆伝播する。

本研究で用いる D4E100P1maxn エージェントでは、ルートノードの直下に、パスを含むその盤面のすべての合法手に対応する手ノードを生成する。これら深さ 1 の手ノードについては枝刈りを行わず、すべてを探索対象として保持する。各深さ 1 の手ノードを起点として、その下に生成されたノードの UCT 値を比較し、UCT 値が最大となる分岐を優先してプレイアウトを行う。葉ノードにおけるプレイアウト回数が 100 回に達した時点でそのノードを展開する。ただし、探索木の展開は深さ 4 までとし、それを超える展開は行わない。この最大探索深度 4 の設定を D4、展開に必要なプレイアウト回数 100 の設定を E100 と表している。

2.2 多人数確率手番ゲームへの適用

アタック 25 は 4 人ゲームであるため、本エージェントでは各盤面の評価値を、4 人のプレイヤーそれぞれの推定勝率からなる 4 次元ベクトル

$$\mathbf{V} = (V_R, V_G, V_W, V_B)$$

として保持する。

また、本研究では、次にどのプレイヤーが手番を取得するかが確率的に決定される確率手番ゲームとしてアタック 25 をモデル化している。このため、探索木において各手ノードの下に、赤・緑・白・青の 4 人それぞれに対応する 4 つのチャンスノードを配置する。プレイヤー i に対応するチャンスノードの下には、そのプレイヤーが手番を取得した場合に選択可能な合法手に対応する手ノードを配置する。

したがって、探索木は模式的には図 1 のような構造を持つ。後述する探索幅の制限は、この各チャンスノードの下に生成される合法手を対象として行う。

プレイアウトによって得られた勝敗結果は、4 次元の評価ベクトルとして探索木を逆伝播する。合法手に対応する子ノードからチャンスノードへの逆伝播については、通常の UCT と同様に、プレイアウト回数と評価値を更新する。一方、4 つのチャンスノードからその親である手ノードへの逆伝播では、各プレイヤーが次の手番を取得する確率を用いて期待値を計算する。

赤・緑・白・青の手番獲得確率をそれぞれ p_R, p_G, p_W, p_B とし（ただし $p_R + p_G + p_W + p_B = 1$）、各プレイヤーに対応するチャンスノードの評価ベクトルを

$$\mathbf{V}_R, \quad \mathbf{V}_G, \quad \mathbf{V}_W, \quad \mathbf{V}_B$$

© 2026 Information Processing Society of Japan

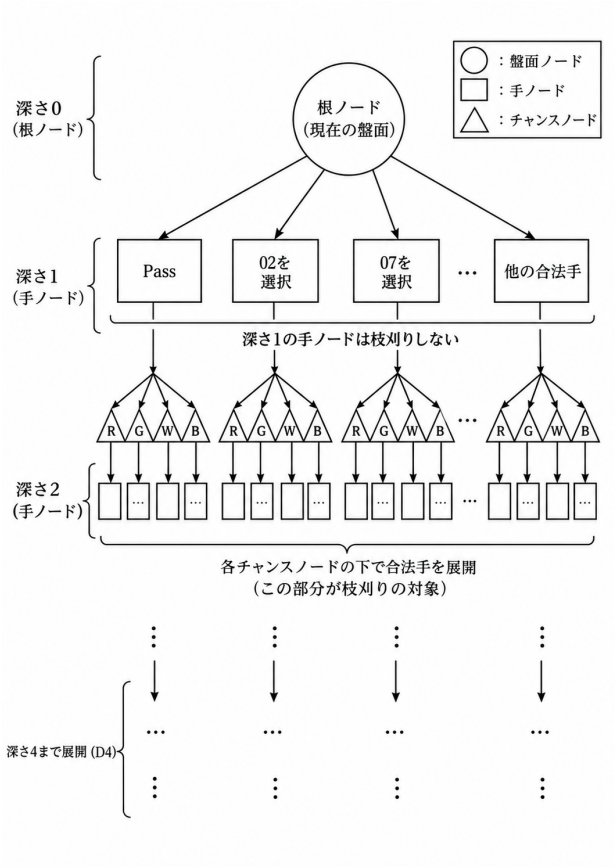


図 1　探索木の模式図。

とするとき、親ノードに格納する評価値を

$$\mathbf{V} = p_R\mathbf{V}_R + p_G\mathbf{V}_G + p_W\mathbf{V}_W + p_B\mathbf{V}_B$$

とする。特に手番獲得確率を指定しない場合には、各プレイヤーが等確率で手番を取得するものとし、

$$p_R = p_G = p_W = p_B = 0.25$$

に設定している。

この構造により、各プレイヤーが自己の推定勝率を評価しながら行動を選択する多人数ゲームの性質と、次の手番者が確率的に決定される確率手番ゲームの性質を、同一の探索木の中で扱うことができる。

2.3　探索幅を制限した UCT

標準的な UCT では、一度生成された複数の手ノードを探索対象として保持し、UCT 値に基づいて探索資源を配分する。一方、本研究の探索木では、一つの手ノードの下に 4 つのチャンスノードがあり、さらに各チャンスノードの下に複数の合法手ノードが生成されるため、分岐数が大きくなりやすい。このため、限られた計算資源の下ですべての分岐を保持すると、個々の分岐に十分な探索回数を割り当てることが難しく、探索深度を確保しにくい。

そこで本研究で用いるエージェントでは、各合法手についてまず少数回の初期プレイアウトを行い、その後

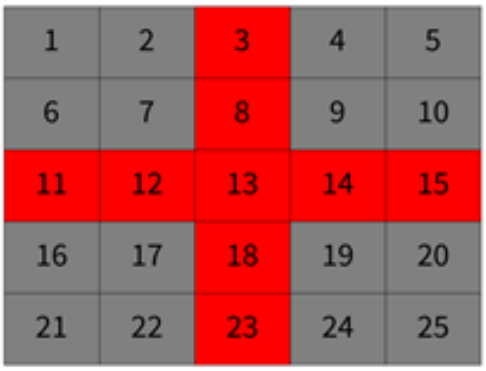


図 2　単色十字型盤面。

UCT 値に基づいて探索資源を逐次的に配分する。本実装では、各合法手に 10 回の初期プレイアウトを割り当てる。その後、ある手ノードのプレイアウト回数が所定の展開条件に最初に到達した時点で、当該手ノードを残し、他の手ノードを探索対象から除外する。

本研究で用いる P1 設定では、深さ 1 の手ノードについてはすべて保持する一方、深さ 2 以降では各チャンスノードの下で探索幅を制限する。ここでは、各チャンスノードの下で 1 つの手ノードのみを保持する。この処理は標準 UCT に含まれるものではなく、探索幅を制限することによって、限られた計算資源の下でより深い盤面まで探索するために導入したヒューリスティックである。

この枝制限では有限回のプレイアウト結果によって残される枝が決定されるため、評価値には確率的な変動が生じ得る。この影響を評価するため、本研究では同一条件で探索を繰り返し、評価値の平均および標準偏差を調べる。

2.4　maxN による行動選択

多人数ゲームの探索法として、本シミュレータでは maxN と paranoid のいずれの評価法が利用可能である。

maxN では、各プレイヤーの手番において、そのプレイヤー自身の推定勝率を最大化する手を選択する。これに対し paranoid では、ルートプレイヤー以外のプレイヤーが、ルートプレイヤーの評価値を最小化するように行動すると仮定する。すなわち、ルートプレイヤー対その他 3 人という実質的な二陣営ゲームとして探索を行う。

本研究の目的は、4 人のプレイヤーがそれぞれ自己の推定勝率を最大化しようとしたときに、誰も自発的に手番を取得しようとしない盤面が生じるかを調べることである。したがって、他の 3 人が協調して特定のプレイヤーを不利にすることを仮定する paranoid より、各プレイヤーがそれぞれ自己の評価値を最大化する maxN の方が本研究のゲームモデルに適していると判断し、maxN を採用した。

© 2026 Information Processing Society of Japan

3. 対象盤面と評価方法

3.1 対象とする単色十字型盤面

本研究では、図 2 に示す単色十字型盤面を対象とする。この盤面では、赤のプレイヤーが

$$3, 8, 11, 12, 13, 14, 15, 18, 23$$

の 9 枚のパネルを所有し、他のパネルは未取得である。以下、この盤面を s と表す。図中のパネル番号は通常の整数表記とし、1 桁の番号には先頭の 0 を付さない。一方、着手を記号で表す場合にはパネル番号を 2 桁で表し、例えば赤のプレイヤーが 2 番のパネルを選択する着手を $R02$、緑のプレイヤーが 7 番のパネルを選択する着手を $G07$ と記す。

3.2 盤面評価と対称性

以下、プレイヤー i に対する盤面 s の推定勝率を $V_i(s)$ と表す。$V_i(s)$ は、前章で述べた UCT エージェントによって得られるプレイヤー i の推定勝率である。本実験では、探索に含まれる確率的変動の影響を軽減するため、同一盤面について 100 回のシミュレーションを行い、その平均値を $V_i(s)$ として用いる。

図 2 の盤面 s は、盤面中央を中心とする 90 度回転、ならびに水平・垂直方向のに対して不変である。このため、この盤面から選択可能なパネルは角（四隅）の $1, 5, 21, 25$ を除く 12 か所であるが、盤面の対称性により、これらは

$$\{2, 4, 6, 10, 16, 20, 22, 24\} \quad と \quad \{7, 9, 17, 19\}$$

の 2 つの同値類に分かれる。したがって、あるプレイヤーが手番を得た場合のすべての合法手を評価するためには、各同値類の代表として 2 番のパネルおよび 7 番のパネルを選択した場合を調べれば十分である。

例えば赤のプレイヤーについて、ゲームの対称性から

$$V_R(s_{R02}) = V_R(s_{R04}) = \cdots = V_R(s_{R24})$$

および

$$V_R(s_{R07}) = V_R(s_{R09}) = V_R(s_{R17}) = V_R(s_{R19})$$

が期待される。有限回のシミュレーションではこれらの推定値は完全には一致しないが、本研究では対称な着手についても一部を実際に評価し、得られる値が対称性と整合していることを確認した。

また、盤面 s では緑・白・青の 3 プレイヤーはいずれもパネルを所有していないため、盤面およびゲームルールに関して対称な立場にある。したがって理論上

$$V_G(s) = V_W(s) = V_B(s)$$

であり、対応する着手後盤面についても同様の対称性が成り立つ。

4. 全員パス戦略のナッシュ均衡

4.1 同時手番ゲームとしての定式化

前節で述べたように、本研究ではクイズが出題された時点において、各プレイヤーが「パスする」または「解答する」のいずれかを同時に選択するゲームを考える。解答を選択したプレイヤーが 1 名以上存在する場合、その中から等確率で 1 名が手番を取得するものとする。一方、全プレイヤーがパスを選択した場合には盤面は変化せず、次のクイズへ移行するものとする。

全プレイヤーがパスを選択する戦略プロファイルを (P, P, P, P) とする。このとき盤面は変化しないため、プレイヤー i の利得を $u_i(P, P, P, P) = V_i(s)$ とする。

次に、他の 3 プレイヤーがパスを選択しているときに、プレイヤー i のみが解答を選択する場合を考える。この場合、プレイヤー i は確率 1 で手番を取得する。盤面 s におけるプレイヤー i の合法手の集合を $A_i(s)$、合法手 a を選択した後の盤面を s_a とする。このとき、解答を選択した場合に得られる最大の評価値は

$$\max_{a \in A_i(s)} V_i(s_a)$$

である。したがって、

$$V_i(s) > \max_{a \in A_i(s)} V_i(s_a) \tag{4.1}$$

がすべてのプレイヤー i について成り立てば、他の 3 プレイヤーがパスを選択しているとき、プレイヤー i にとってもパスが最適反応となる。

4.2 推定勝率による最適反応の比較

各盤面について 100 回のシミュレーションを行った結果を以下の表に示す。括弧内は 100 回の推定値の標準偏差である。

プレイヤー R（赤）

パス $V_R(s)$	0.445600 (0.007771)
2 番のパネルを選択後	0.430471 (0.011144)
7 番のパネルを選択後	0.406977 (0.017239)
解答時の最大値	0.430471
パスとの差	0.015129

プレイヤー G（緑）

パス $V_G(s)$	0.185353 (0.003298)
2 番のパネルを選択後	0.142092 (0.003590)
7 番のパネルを選択後	0.149361 (0.005863)
解答時の最大値	0.149361
パスとの差	0.035992

プレイヤー W（白）

© 2026 Information Processing Society of Japan

パス $V_W(s)$	0.184561 (0.003421)
2 番のパネルを選択後	0.141226 (0.004207)
7 番のパネルを選択後	0.149698 (0.006192)
解答時の最大値	0.149698
パスとの差	0.034863

プレイヤー B（青）

パス $V_B(s)$	0.184486 (0.003662)
2 番のパネルを選択後	0.141968 (0.003962)
7 番のパネルを選択後	0.149307 (0.005538)
解答時の最大値	0.149307
パスとの差	0.035179

赤のプレイヤーについては 2 番のパネル選択が 7 番のパネル選択より高い評価値を与えるが、

$$V_R(s) = 0.445600 > V_R(s_{R02}) = 0.430471$$

である。したがって、赤が取り得るすべての合法手について

$$V_R(s) > V_R(s_a) \qquad (a \in A_R(s))$$

が成り立つ。

一方、緑・白・青については 7 番のパネル選択が 2 番のパネル選択より高い評価値を与える。しかし、それぞれ

$$V_G(s) = 0.185353 > V_G(s_{G07}) = 0.149361,$$
$$V_W(s) = 0.184561 > V_W(s_{W07}) = 0.149698,$$
$$V_B(s) = 0.184486 > V_B(s_{B07}) = 0.149307$$

であり、いずれについても解答して最善の合法手を選択した場合より、パスした場合の評価値の方が高い。

以上より、すべてのプレイヤー i に対して式 (4.1) が成立する。したがって、他の 3 プレイヤーがパスを選択しているとき、各プレイヤーにとってパスは最適反応である。よって、100 回のシミュレーション平均によって構成された推定利得ゲームにおいて、(P, P, P, P) はナッシュ均衡である。さらに、すべてのプレイヤーについて式 (4.1) が狭義の不等式として成立していることから、本モデルにおいて (P, P, P, P) は狭義ナッシュ均衡（strict Nash equilibrium）である。

なお、ここで主張しているのは、他のプレイヤーがパスを選択しているときにパスが最適反応となることであり、パスが他のプレイヤーの任意の戦略に対して常に最適となる支配戦略であることを意味しない。

4.3 Regret Matching との関係

以上の議論から、全員パスがナッシュ均衡であることは Regret Matching の計算とは独立に確認できる。次節では、Pass/Answer の全 16 通りの戦略プロファイルに対して推定利得を与え、Regret Matching を適用することで、相関均衡の観点からこの盤面の戦略構造をさらに調べる。

特に、純粋戦略ナッシュ均衡は一点に集中した相関均衡でもあるため、本節の結果から (P, P, P, P) に確率 1 を与える分布は相関均衡である。次節の Regret Matching の計算は、この均衡の存在を証明するためではなく、学習過程から得られる経験分布が実際にこの均衡へ集中するかを調べるものと位置づける。

5. リグレット・マッチングによる相関均衡解析

5.1 Pass / Answer ゲームの利得表

前節では、他の 3 プレイヤーが Pass を選択した場合の各プレイヤーの最適反応を調べ、全員が Pass を選択する戦略プロファイル (P, P, P, P) が狭義ナッシュ均衡となることを示した。本節では、各プレイヤーが Pass または Answer を同時に選択する 16 通りの戦略プロファイル全体を対象として、相関均衡の観点から戦略構造を解析する。

プレイヤーの順序を R, G, W, B とする。前節と同様に、各盤面について D4E100P1maxn エージェントによる評価を乱数系列を変えて 100 回実行し、得られた推定勝率の平均値をその盤面の評価値として用いる。元盤面 s における評価値ベクトルを

$$\mathbf{V}_0 = (0.445600,\ 0.184800,\ 0.184800,\ 0.184800)$$

とした。ここで緑・白・青は盤面およびゲームルールに関して対称であるため、100 回のシミュレーションから得られた各評価値を対称化して用いた。

また、各プレイヤーが単独で Answer を選択し、手番取得後に自己の推定勝率を最大化する着手を選んだ場合の評価値ベクトルをそれぞれ

$$\mathbf{U}_R = (0.430471,\ 0.189843,\ 0.189843,\ 0.189843),$$
$$\mathbf{U}_G = (0.407598,\ 0.149361,\ 0.221521,\ 0.221521),$$
$$\mathbf{U}_W = (0.407598,\ 0.221521,\ 0.149361,\ 0.221521),$$
$$\mathbf{U}_B = (0.407598,\ 0.221521,\ 0.221521,\ 0.149361)$$

とした。それぞれの値は、対応する着手後盤面を 100 回評価した平均値を用い、ゲームルールに関して対称となる値については対称化して用いている。赤のプレイヤーについては 2 番のパネルを選択した場合が最も高い評価値を与え、緑・白・青の 3 プレイヤーについては 7 番のパネルを選択した場合が最も高い評価値を与える。ここで 2 番および 7 番は、それぞれ対称性によって同値となる合法手の代表である。

Answer を選択したプレイヤーの集合を S とする。$S \neq \varnothing$ の場合には、集合 S に属する各プレイヤーが等確率で手番を取得するものとしているため、戦略プロファイルに対する利得ベクトルを

© 2026 Information Processing Society of Japan

$$\mathbf{u}(S) = \frac{1}{|S|} \sum_{i \in S} \mathbf{U}_i \tag{5.1}$$

によって定めた。一方、全員が Pass を選択した場合には、

$$\mathbf{u}(\varnothing) = \mathbf{V}_0 \tag{5.2}$$

とした。これにより、16 通りすべての純粋戦略プロファイルに対する 4 人の利得が定まる。実装においても、Answer を選んだプレイヤーに対応する単独着手後の評価ベクトルを平均することで 16 通りの利得表を生成した。

5.2 利得表の構造

以下、a_{-i} はプレイヤー i 以外の 3 人の行動の組を表す。

この利得表には、単に (P,P,P,P) がナッシュ均衡であるという以上の特徴がある。緑のプレイヤーについて、他の 3 プレイヤーの行動を固定し、Pass と Answer の利得を比較すると、$u_G(P, a_{-G}) > u_G(A, a_{-G})$ がすべての a_{-G} について成立する。実際、その差の最小値は

$$\min_{a_{-G}} \{u_G(P, a_{-G}) - u_G(A, a_{-G})\} = 0.015400 > 0$$

である。白・青についても対称性から同様である。したがって、緑・白・青の 3 プレイヤーにとって Pass は Answer を狭義に支配する行動である。

相関均衡において、狭義に支配された行動が正の確率で推薦されることはない。したがって、任意の相関均衡において緑・白・青は確率 1 で Pass を選択する。この条件のもとで赤について比較すると、

$$u_R(P,P,P,P) = 0.445600$$

に対し、

$$u_R(A,P,P,P) = 0.430471$$

であるから、赤についても Pass が狭義の最適反応となる。したがって、本研究で構成した推定利得ゲームにおける相関均衡は (P,P,P,P) に確率 1 を与える分布のみであり、これは唯一の相関均衡である。

以下では、この相関均衡に対して Regret Matching による学習過程がどのように振る舞うかを数値的に検証する。

5.3 Regret Matching

Hart ら [5] の Regret Matching を用いる。時刻 t におけるプレイヤー i の行動を $a_i^t \in \{P, A\}$、4 人の行動プロファイルを

$$a^t = (a_R^t, a_G^t, a_W^t, a_B^t)$$

とする。

プレイヤー i が行動 j を実際に選択した盤面について、「そのとき行動 k を選択していたならばどれだけ利得が増加したか」を累積する。時刻 T までの j から k への conditional regret を

$$R_i^T(j,k) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}_{\{a_i^t = j\}} \left[u_i(k, a_{-i}^t) - u_i(j, a_{-i}^t) \right] \tag{5.3}$$

と定義する。ここで $\mathbf{1}$ は指示関数である。また、

$$[R_i^T(j,k)]_+ = \max\{R_i^T(j,k), 0\}$$

とする。

次の反復において、現在の行動が j であるとき、別の行動 k へ変更する確率を

$$\Pr(j \to k) = \frac{[R_i^T(j,k)]_+}{\mu} \tag{5.4}$$

とした。本研究では各プレイヤーが 2 行動のみを持ち、すべての利得が 0 以上 1 以下であるため、正規化定数として $\mu = 1$ を用いた。行動を変更しない確率は式 (5.4) の余事象として定める。

以上の更新を各プレイヤーについて繰り返す。実装では、実際に選択された行動 j から他方の行動 k への反実仮想的な利得差のみを累積し、その正の平均 regret に比例して次の行動への遷移確率を定めている。

5.4 収束の評価

Regret Matching では、各時点における行動確率そのものではなく、これまでに実現した行動プロファイルの経験分布が相関均衡へ近づく [5]。そこで本研究では、時刻 T までに (P,P,P,P) が出現した累積頻度

$$q_T(PPPP) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}_{\{a^t = (P,P,P,P)\}} \tag{5.5}$$

を第 1 の収束指標とした。

また、相関均衡条件からの乖離を示す指標として、

$$R_{\max}(T) = \max_{i,\, j \neq k} [R_i^T(j,k)]_+ \tag{5.6}$$

を用いた。すべての正の conditional regret が 0 へ近づけば、過去の行動を別の行動へ置き換えることで利得を改善できる誘因が消失することを意味する。実装では式 (5.5) と式 (5.6) を一定反復回数ごとに記録した。

本実験では初期戦略プロファイルを (A,A,A,A) とし、100,000 回の反復を行った。

5.5 シミュレーション結果

図 3 に、全員 Pass の累積経験頻度 $q_T(PPPP)$ の推移を示す。反復の初期には複数の戦略プロファイルが出現するが、反復回数の増加とともに $q_T(PPPP)$ は急速に 1 へ近づいた。100,000 反復後には

$$q_T(PPPP) = 0.998110$$

© 2026 Information Processing Society of Japan

となり、経験分布のほぼすべてが全員 Pass の戦略プロファイルに集中した。

図 4 に、最大の正の conditional regret $R_{\max}(T)$ の推移を示す。反復初期には正の regret が存在するものの、反復とともに急速に減少し、100,000 反復後には

$$R_{\max}(T) = 1.96 \times 10^{-5}$$

となった。

したがって、本実験では

$$q_T(PPPP) \longrightarrow 1, \qquad R_{\max}(T) \longrightarrow 0$$

という挙動が観測された。これは、前節までの利得解析によって得られた唯一の相関均衡 (P, P, P, P) に、Regret Matching の経験分布が集中することと整合している。

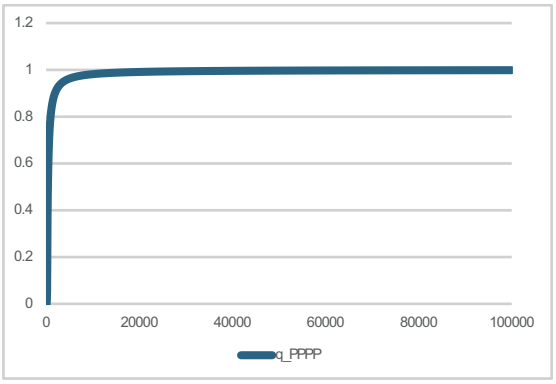


図 3 全員が Pass を選択する戦略プロファイル (P, P, P, P) の累積頻度 $q_T(PPPP)$。

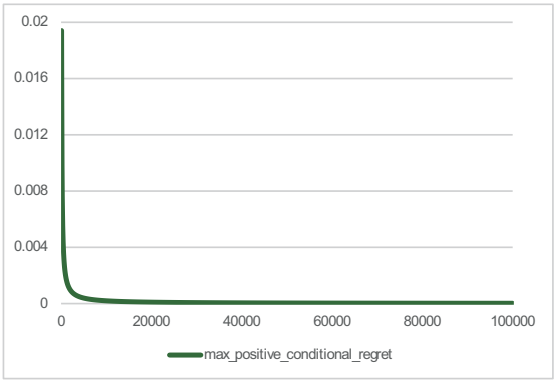


図 4 最大の正の conditional regret $R_{\max}(T)$。

6. 結論

本研究では、アタック 25 を、各プレイヤーが Pass または Answer を同時に選択し、Answer を選択したプレイヤーの中から確率的に手番者が決定されるゲームとしてモデル化した。盤面評価には、各盤面を 100 回評価した推定勝率の平均値を用いた。単色十字型盤面について UCT エージェントによる盤面評価を行った結果、各プレイヤーにとって、他の 3 人が Pass を選択しているときには、自らも Pass を選択することが最適反応となることを確認した。したがって、全員 Pass の戦略プロファイルは狭義ナッシュ均衡である。さらに、Pass/Answer の 16 通りの戦略プロファイルから構成される推定利得ゲームを解析し、全員 Pass が唯一の相関均衡となることを示した。Regret Matching による学習シミュレーションにおいても、経験分布はこの均衡に集中し、最大の正の conditional regret は 0 に近づいた。以上から、アタック 25 には、本論文で定義する意味での手詰まり盤面が実際に存在することが示された。

7. 今後の展開

7.1 他の手詰まり盤面の探索と特徴付け

本研究では、単色十字型盤面について、全プレイヤーが Pass を選択する戦略プロファイルが狭義ナッシュ均衡となることを示した。今後は、他の盤面についても、各プレイヤー i に対して式 (4.1)

$$V_i(s) > \max_{a \in A_i(s)} V_i(s_a)$$

が成立するかを系統的に調べ、手詰まり盤面を探索する。さらに、発見された盤面について、パネル配置、獲得枚数、対称性などの特徴を比較することで、手詰まり盤面が生じるための十分条件や特徴的構造を明らかにすることを課題とする。手詰まり盤面を単に列挙するだけでなく、どのような盤面で式 (4.1) が成立しやすいかを特徴付けることができれば、盤面構造と戦略均衡との関係を明らかにできる。

7.2 停滞コストを導入した動的モデル

本研究では、各プレイヤーの勝率のみを利得として扱ったため、全員 Pass が狭義ナッシュ均衡となった。一方、実際の番組では全員が Pass を選択した場合にも次の問題へ進み、ゲームは継続する。そこで、全員 Pass による停滞に小さなコストを導入すると、「誰が解答してゲームを進行させるか」という volunteer's dilemma 型の問題が生じると考えられる。解答による勝率低下と停滞コストの関係を導入することで、手詰まり状態がどのように解消されるかを解析することも今後の課題である。本研究の単色十字型盤面では、解答による推定勝率の低下量は赤のプレイヤーが最も小さいため、停滞コストを導入した場合には赤が最初に Answer を選択する可能性がある。

7.3 非一様な手番獲得確率への拡張

本研究では、特に指定しない場合、各プレイヤーの手番獲得確率を

$$p_R = p_G = p_W = p_B = 0.25$$

としている。しかし、実際の番組では、クイズの得意不

© 2026 Information Processing Society of Japan

得意などにより、手番獲得確率はプレイヤーごとに異なると考えられる。そこで、

$$(p_R, p_G, p_W, p_B), \qquad p_R + p_G + p_W + p_B = 1$$

を一般に変化させた場合に、手詰まり盤面や全員 Pass の均衡がどの範囲で維持されるかを調べることが考えられる。これは、確率手番ゲームとしてのアタック 25 の性質をさらに明らかにする上で重要な拡張である。

参考文献

[1] 阿原一志, 浜向直, “アタック 25 の UCT エージェントとある定跡について”, GPW2025 (2025).

[2] 野口陽来, 松井利樹, 小森成貴, 橋本隼一, 橋本剛, “アタック 25 の最適戦略”, GPW2007 (2007).

[3] 阿原一志, 浜向直, “A25platform AI 研究のためのアタック 25 シミュレータ” (2025). `https://github.com/aharalabMeiji/A25_platform`

[4] Kocsis L. and Szepesvári C., “Bandit based Monte-Carlo Planning”, Machine Learning: ECML 2006 (2006).

[5] Hart, S. and Mas-Colell, A.,“A Simple Adaptive Procedure Leading to Correlated Equilibrium”, Econometrica (2000).

© 2026 Information Processing Society of Japan