

Title	不完全情報同時手番ゲームにおいて人間のような読み合いを演出するAIプレイヤー
Author(s)	小西, 健太郎; シュエ, ジュウシュエン; 池田, 心
Citation	情報処理学会 第49回GI研究発表会, 2023-3
Issue Date	2023-03-17
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/18226
Rights	社団法人 情報処理学会, 小西 健太郎, シュエ ジュウシュエン, 池田 心, 情報処理学会 第49回GI研究発表会, 2023-3, 2023. ここに掲載した著作物の利用に関する注意 本著作物の著作権は情報処理学会に帰属します。本著作物は著作権者である情報処理学会の許可のもとに掲載するものです。ご利用に当たっては「著作権法」ならびに「情報処理学会倫理綱領」に従うことをお願いいたします。 Notice for the use of this material: The copyright of this material is retained by the Information Processing Society of Japan (IPSJ). This material is published on this web site with the agreement of the author (s) and the IPSJ. Please be complied with Copyright Law of Japan and the Code of Ethics of the IPSJ if any users wish to reproduce, make derivative work, distribute or make available to the public any part or whole thereof. All Rights Reserved, Copyright (C) Information Processing Society of Japan.
Description	情報処理学会 第49回GI研究発表会, 2023-3, 明治大学

不完全情報同時手番ゲームにおいて人間のような読み合いを演出する AI プレイヤ

小西 健太郎^{1,a)} 池田 心^{1,b)} シュエ ジュウ シュエン^{1,c)}

概要：不完全情報ゲームにおいて、人間プレイヤがナッシュ均衡戦略を計算することは現実的に難しいため、予測される相手の情報や行動に対してアドバンテージを得られる行動を選択する「読み」を用いて意思決定を行う場合がある。人間プレイヤ同士が読みの思考に基づいて意思決定を行う場合や、相手が読んでくことを考慮して意思決定を行う「読み合い」は、一部のゲームでは主流な意思決定方法の1つであり、この特有の駆け引きはゲームを楽しむ上で重要な要素の1つといえる。本研究では、ポケモンバトルを簡略化したゲーム上で、人間のような読み合いを演出する AI プレイヤを提案した。読み合い演出の前提として、人間が自然に感じやすい着手戦略である、 δ -ナッシュ均衡戦略というノイズを加えた利得行列を用いる手法の有効性を示した。また、 δ -ナッシュ均衡戦略に人間プレイヤのような癖や傾向をバイアスとして付与し、多様な着手戦略が得られることを確認した。さらに、人間の読みを模倣した着手として、相手モデリングによって次の行動を予測し、搾取戦略を生成する手法についてアプローチを示した。

キーワード：同時手番ゲーム、ポケモン、読み合い、ナッシュ均衡戦略

1. はじめに

近年、人工知能技術 (AI) は急速に発達しており、囲碁における AlphaGo[1] をはじめとして、様々なゲームで AI プレイヤが人間のトッププレイヤ以上の水準に到達しつつある。そのため、新たな方向性として、「楽しませる」「教える」など AI プレイヤに強い対戦相手以外の多様な用途を持たせる研究 [2], [3] が盛んに行われているものの、課題もまだ多い。

プレイヤが一部の情報を正確に把握できない不完全情報ゲームには、「付け込まれない戦略」であるナッシュ均衡戦略が存在するが、複雑なゲームで人間プレイヤがこの戦略を求めることは現実的に難しい。そのため人間プレイヤは、相手の情報や行動を予測し、予測に対してアドバンテージを得られる行動を選択する「読み」を用いる場合がある。人間同士が読みの思考に基づいて意思決定を行う場合や、相手が読んでくことを考慮して意思決定を行う「読み合い」は一部のゲームでは主流な意思決定方法の1つであり、この特有の駆け引きはゲームを楽しむ上で重要な要素の1

つといえる。

人間の「読み」の思考に近い手法としては、相手プレイヤをモデリングすることで、次の行動などを予測する相手モデルが研究されている。しかし、「読み合い」の駆け引きに着目した AI プレイヤは少なく、人間プレイヤの対戦相手として読み合いの駆け引きを演出することは難しい。この点は一部の不完全情報ゲームにおける AI プレイヤの課題の1つだと考える。

そこで、本研究では不完全情報ゲームの一種である同時手番ゲームにおいて、AI プレイヤに人間のような読み合いを演出させる手法の確立を目的とする。読み合いを演出する手法として、対象ゲームにおいて人間が自然に感じる着手戦略を生成し、人間のような癖や傾向を持つ着手、人間の読みを模倣する着手を行うことで実現を目指す。

2. 関連研究

不完全情報ゲームにおいてナッシュ均衡戦略はゲームの解であるものの、複雑なゲームで均衡戦略を求めることは容易ではない。複雑な不完全情報ゲームや一部のリアルタイムゲームにおいて、高い利得が得られる戦略を獲得する手法として相手モデル (opponent model) がある。相手モデルは、相手プレイヤをモデリングすることで、相手の傾向から次の行動などを予測する手法である。

¹ 北陸先端科学技術大学院大学
Japan Advanced Institute of Science and Technology, Nomi,
Ishikawa 923-1211, Japan
^{a)} s2110076@jaist.ac.jp
^{b)} kokolo@jaist.ac.jp
^{c)} hsuehch@jaist.ac.jp

Van der Kleij は、ポーカーの 1 種である Texas Hold'em においてプレイスタイルに基づいてプレイヤをクラスタリングし、それぞれのクラスタに対応するモデルを用いることでプレイヤの行動などを予測する手法を提案している [4].

また、Sato らは、格闘ゲームにおいて複数のルールベースプレイヤを組み合わせ、相手に応じてプレイヤを切り替える手法を提案している [5]. この手法ではプレイヤの切り替えを、与えたダメージと受けたダメージの差分を報酬とする多腕バンディット問題と捉えることで、相手に有利な行動を取れるよう制御している.

3. 対象ゲーム

人間プレイヤ間で読み合いが生じやすい同時手番ゲームとして代表的なものにポケモンバトルがある. 最も一般的なシングルバトルルールでは、2 人のプレイヤが 1 匹ずつ場にポケモンを出し、1 ターンに一度、攻撃/交換などのコマンド選択を同時手番で行う. また、ポケモンにはタイプと呼ばれる相性関係があり、人間プレイヤ間ではタイプ相性をもとに相手プレイヤの行動を読み合う戦略は主流な意思決定方法である.

しかし、ポケモンバトルはパーティーの作成やポケモンの選出など多段階の要素で構成されていることや、ポケモンの種類やワザ、ステータスなど高すぎる自由度や不確実性の問題から、ゲーム AI 研究において非常に難易度の高い課題の 1 つといえる. そのため、本研究ではポケモンバトルをもとにして、読み合いが生じるような本質は残しつつも、高度または些末な要素を取り除いた簡易版ポケモンバトルを作成した.

3.1 簡易版ポケモンバトル

以下に簡易版ポケモンバトルの詳細なルールについて示す.

- (1) 概要: 簡易版ポケモンバトルは、お互いのプレイヤが場に 1 匹ずつポケモンを出し、1 ターンに 1 度コマンドを選択するシングルバトル形式である.
- (2) 用いることができるポケモン: 簡易版ポケモンバトルで用いるポケモンは以下の要素で構成されている.
 - (a) ステータス: ポケモンは HP、すばやきの 2 種類のステータスを持つ. HP の最大値は 5, すばやきの最大値は 10 となっている.
 - (b) タイプ: ポケモンはほのお, みず, くさ, でんきの 4 種類のタイプのいずれかをもつ. 各タイプの相性関係は表 1 を参照.
 - (c) ワザ: 各ポケモンは 3 つのワザを持つ. ワザにはそれぞれタイプと威力が存在する.
- (3) ゲーム内で可能な行動: ゲーム内で選択可能なコマンドは以下の 2 種類である.

- (a) ワザ (攻撃): ワザは相手ポケモンに攻撃する行動であり、タイプと威力の 2 種類のパラメータがある. 威力は基本 2 ダメージだが、攻撃を受ける(ぼうぎょ) ポケモンのタイプと、ワザのタイプの相性関係によって威力が変化する. 基本ダメージに表 1 の倍率をかけた値が、最終的に相手に与えるダメージになる. また、補正倍率が 1.5 のとき、さらに攻撃するポケモンとワザのタイプが一致した場合には倍率が 2.0 となる.
- (b) 交換: 交換は場に出ているポケモンと、控えの HP が 0 でないポケモンを入れ替える行動である. 交換を行ったターンは、新たに場に出たポケモンはそれ以上行動選択ができない.
- (4) ゲームの進行: ゲームは以下に示す手順で進行される.
 - (a) パーティーの決定: ゲームが開始するとプレイヤには表 2 に示す 2 種類のパーティーのうち、どちらかのパーティーがランダムに与えられる. プレイヤは相手のポケモンの種族と情報 (HP, すばやき, ワザの種類) を確認することができる.
 - (b) 場に出るポケモンの決定: ゲーム開始時にお互いのパーティーから、それぞれ 1 匹ずつランダムに選ばれたポケモンが場に出される.
 - (c) コマンド選択: 各プレイヤは 1 ターンに一度、自分の場に出ているポケモンの持つワザ 3 種と、控えのポケモンへの交換 2 種の最大 5 種類の選択肢から 1 つを同時に選択する. 選択したコマンドの実行は、交換、すばやきの高いポケモンのワザ、すばやきの低いポケモンのワザの順で処理される*1. また、選択したコマンドが実行される前に自分の場に出ているポケモンの HP が 0 になった場合には、選択していた行動はキャンセルされ、まだ HP が 0 でないポケモンがいる場合、それらのポケモンからプレイヤが選択して場に出す.
 - (d) 勝敗の判定: 相手プレイヤのパーティーのポケモンすべての HP を 0 にしたプレイヤの勝利となる. また、20 ターン目をむかえた時点でゲームの勝敗がついていない場合は、以下に示す順で勝敗を判定する.
 - (i) 20 ターン目時点で、残りポケモン数 (HP が 0 でないポケモン) が多いプレイヤの勝利となる.
 - (ii) 残りポケモン数が同じ場合、残りポケモンの HP 総数が大きいプレイヤの勝利となる.
 - (iii) HP 総数が同じ場合引き分けとなる.

*1 簡易版ポケモンバトルではすばやきが同じポケモンがないよう設定した.

表 1 各タイプの相性関係と、ダメージの補正倍率

こうげき	ぼうぎょ				
	ほのお	くさ	でんき	みず	
ほのお	0.5	1.5	1.0	0.5	
くさ	0.5	0.5	1.5	1.0	
でんき	1.0	0.5	0.5	1.5	
みず	1.5	1.0	0.5	0.5	

表 2 パーティの各パラメータ

パーティー	パーティー 1			パーティー 2		
ポケモン	A	B	C	D	E	F
タイプ	くさ	みず	でんき	くさ	ほのお	でんき
HP	5	5	5	5	4	4
すばやさ	5	4	10	6	7	8
覚えるワザ	くさ	みず	でんき	くさ	ほのお	でんき
	ほのお	ほのお	くさ	ほのお	みず	くさ
	でんき	でんき	みず	でんき	くさ	みず

4. 提案手法

不完全情報ゲームにおいて人間プレイヤーが読み合いが起きていると感じ取る場面は実に様々ではあるが、代表的な状況としては「相手の情報や行動を読んでアドバンテージを得ることができたとき」、「相手に自身の情報や行動を読まれたと思ったとき」の 2 つが挙げられる。そのため、対戦中にこのような体験をすることができれば人間プレイヤーは読み合いが起きていることを実感することができると考え、人間のような読み合いを演出する手法を検討した。

本研究では対戦中に前述のような状況をつくる手法として、(1) 人間が自然に感じる着手戦略、(2) 人間的な癖や傾向をもった着手戦略、(3) 人間の読みを模倣した着手戦略を提案する。以下、個別具体的に説明する。

人間が自然に感じる着手戦略

多くの人間プレイヤーはゲームを楽しむという観点においては相手に自分が意図を理解できない手や、悪手に感じるような手を選択されることを嫌う。そのため、後述する 2 つの提案手法、特に癖や傾向をバイアスとして付与する元となる戦略として、人間が不自然に感じない着手戦略の生成について取り組む。

人間的な癖や傾向をもった着手戦略

人間プレイヤーは、ゲームにおいて何らかの基本とする考えや、好みを持っており、これらの傾向はゲーム中に一貫して現れる場合が多い。また、人間がある意図を持って選択した行動にはナッシュ均衡戦略と比較して何らかの偏りがあり付け込まれる隙が生じる。人間同士の対戦では、お互いにこの偏りから生じる隙を狙って読み合いになることも多い。このような読まれる可能性がある傾向や癖を持った着手戦略を用いて、隙をつくることで人間プレイヤーにあえて付け入らせる。こ

のような体験によって「相手の行動を読んでアドバンテージを得ることができた」と思わせることができる。と考える。

人間の読みを模倣した着手戦略

人間プレイヤーは殆どの場合、着手に付け込まれる癖や傾向がある。そこで、相手プレイヤーに付け入るような戦略を用いることで、人間プレイヤーに「相手に自分の行動を読まれた」と思わせることができると考える。このように AI プレイヤー側が読みを模倣することで、擬似的にお互いが「読み」の思考に基づいて意思決定を行う状態を表現する。

4.1 提案 AI プレイヤーの概要

前述した提案手法を実際に実現する AI プレイヤーの構成を図 1 に示す。提案 AI プレイヤーは、標準着手生成部、人間的な癖を付与するバイアス着手生成機能、搾取戦略実施機能で構成される。以下、それぞれの機能について概要を説明する。

(1) 標準着手生成部

標準着手生成部では、簡易版ポケモンバトルを対象として、人間プレイヤーが不自然に感じないような合理性のある戦略を生成する。この戦略は、ナッシュ均衡戦略をベースにした混合戦略（確率分布）であるが、人間プレイヤーが不自然に感じない程度に実力を弱めに調整している。

(2) バイアス着手生成機能

バイアス着手生成機能では、標準着手生成部で生成した戦略をもとにして、人間がもつような癖や傾向をバイアスとして付与した戦略を生成する。このバイアスは、AI プレイヤーの選択時に決められ、数試合一貫して用いる。

(3) 搾取戦略実施機能

搾取戦略実施機能は、現在または過去のゲーム中における相手プレイヤーの行動から、相手プレイヤーをモデリングし、次の行動を予測する相手モデルを持つ。相手モデルは、「交換を嫌う」、「相性の良い攻撃を好む」など既定の着手モデルを複数個持っており、相手のこれまでの行動に最も近い着手モデルが示す戦略を次の相手の着手とみなす。そして、予測される着手に対してアドバンテージを得られるような搾取戦略を生成する。ただし、搾取戦略は毎ターン行うわけではなく、モデルの予測に用いる尤度の閾値などに応じて実施する。

5. 標準着手生成部

5.1 モンテカルロ法

不完全情報ゲームは、 $\alpha\beta$ 探索などの従来のゲーム木探索で着手を決定する手法は効果的ではないため、標準着手戦略の一案としてモンテカルロ法をまず検討した。モンテ

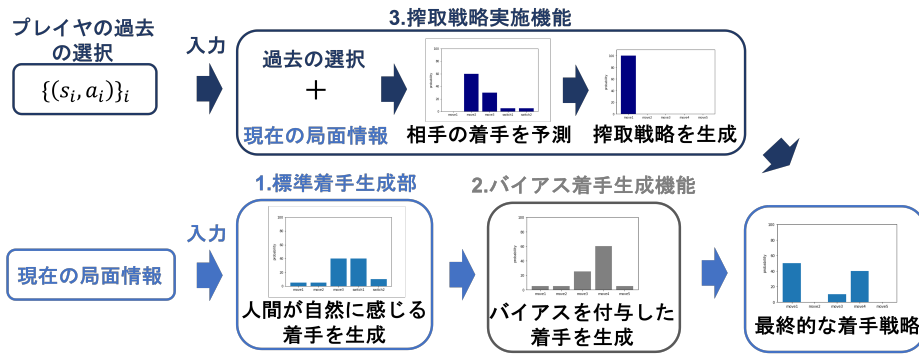


図1 読み合いを演出するAIプレイヤーの構成

カルロ法は、ランダムシミュレーションによるプレイアウトを行うことで状態の良さを推定し、それに基づいて着手を決定する手法である。以下にモンテカルロ法の手順について示す。

- ある状態 s_i において次の行動を決める場合、自身の行動の選択肢5通り $(a_1, \dots, a_5)$ と、相手の行動の選択肢5通り $(b_1, \dots, b_5)$ の組み合わせ計25通りについてランダムシミュレーションを行う。
- 25通りの組み合わせについて、それぞれ20回ずつ勝敗がつくまでシミュレーションを行う。現在の状態 s で行動の組 (a_1, b_1) についてプレイアウトを行う場合、状態 s では (a_1, b_1) を着手し、次状態以降ではそれぞれの行動をランダムで決定しシミュレーションを行う。
- それぞれの組み合わせについてプレイアウトの勝ち数を用いて評価する。 (a_i, b_j) について規定回数のプレイアウトを行ったときの勝ち数を e_{ij} とする。各着手の評価には、行動の組み合わせに対する評価 $(e_{11}, e_{12}, e_{13}, \dots, e_{55})$ までの値を用いる。モンテカルロ法の一般的な作法に従い、着手 a_i の評価は $\sum_j e_{ij}$ とする。
- 各着手の評価をもとに、現在の状態での行動を決定する。

作成した原始モンテカルロプレイヤーと、ランダムプレイヤーを500回対戦させ、その勝率によって性能を評価した。原始モンテカルロプレイヤーは対ランダム:0.972と十分な勝率を示した。続いて、人間が対戦した場合に着手が自然かどうかについて調べた。複数回のテストプレイによって、一部の状態を過大・過小評価してしまうことや、相手が取らないであろう行動のプレイアウト結果を高く評価してしまうことによる不自然な着手が見られた。

5.2 後退解析によるナッシュ均衡戦略

モンテカルロプレイヤーの改善案として、ナッシュ均衡戦略を求めた。ナッシュ均衡戦略はゲームの解であり、付け込まれない戦略であるため、人間が「ひどい手だ」という意味で不自然に感じる着手を行う可能性を減らすことが期

待できる。以下、簡易版ポケモンバトルにおいてナッシュ均衡戦略を求める手法について述べる。

5×5の利得行列で表現できるゲームのナッシュ均衡戦略は、各状態の勝率を求めることができれば線形計画法によって求めることができる。本研究ではナッシュ均衡戦略を求めるにあたり、簡易版ポケモンバトルの状態数がそれほど多くないことや、同時手番以外の不完全情報性や不確実性がないことを利用する。ゲームの全状態を数え、それらの状態について最終ターンから後退解析によって状態を評価することでナッシュ均衡戦略を計算した。以下に後退解析の手順を示す。

(1) 状態数の列挙

まず、ゲームを解析する上で、各ターンにおけるゲームの全状態を列挙する。簡易版ポケモンバトルの全状態数は実際には到達しない状態なども含めると $6^6 \times 3^2 \times 20 = 8,398,080$ である。

(2) 最終ターンにおける状態の評価

簡易版ポケモンバトルではゲームが20ターン目を迎えると、必ず勝敗がつく仕様になっている。そのため、20ターン目時点での状態について、勝ち：勝率1.0、負け：勝率0、引き分け：勝率0.5として、評価を行う。

(3) 最終ターン以外の状態の評価

状態 s で行動 (a_i, b_j) を着手した場合の勝率については、遷移先の状態の評価を確認すれば確定的に求めることができる。そのため、状態 s における各行動の組に対する勝率を最大5×5の利得行列で表すことができる。この利得行列から、線形計画法によってナッシュ均衡戦略を求める。状態 s におけるナッシュ均衡戦略を $(p_{a1}, \dots, p_{a5}), (p_{b1}, \dots, p_{b5})$ 、行動 (a_i, b_j) によって遷移した次状態の勝率を w_{ij} とすると状態 s における真の勝率は $\sum_{i,j} (w_{ij} \cdot p_{ai} \cdot p_{bj})$ となる。

(4) 各ターンにおける全状態の評価

後退解析では、前述の評価方法を用いて最終20ターンから19ターン目と順に各ターンにおけるすべての状態の評価を行う。双方がナッシュ均衡戦略を用いた場合を真の勝率として、全状態の勝率を計算すること

ができる。

(5) ナッシュ均衡戦略の計算

ゲームの全状態の勝率を計算することができれば、任意ターンの任意の状態において、利得行列を作成し、ナッシュ均衡戦略を計算できる*2。

後退解析によってゲームの全状態について勝率による評価を行うことで、ナッシュ均衡戦略に従う AI プレイヤ（以下、ナッシュプレイヤ）を作成した。

ナッシュプレイヤの強さを確認する上で、原始モンテカルロプレイヤおよびランダムプレイヤと 500 回の対戦を行った。勝率は、対ランダム：0.968、対モンテカルロ：0.720 となり、ナッシュプレイヤは、モンテカルロプレイヤにも大きく勝ち越している。本質的にはナッシュ均衡戦略は「付け込まれない戦略」であるが、一方でモンテカルロプレイヤには状態の過大評価・過小評価による不適切な着手があるため、このような結果になったと考える。

さらに、ナッシュプレイヤの着手について人間プレイヤが自然に感じるかどうかについて評価を行った。著者が複数のテストプレイを行った結果、ナッシュプレイヤの着手に不自然に感じる手は少なく、自然な着手の作成を行うことが可能になった。また、ナッシュプレイヤの行う確率的な着手によって、一定程度相手プレイヤに読まれたと感じるような体験を得ることができた。しかし、ナッシュプレイヤはゲームの解の 1 つであり、相手に付け入れられない戦略であることから人間プレイヤの対戦相手としては強すぎるという問題が生じる。また、勝率を少しでも高くするため、人間的には選択してもおかしくない行動を全く選択しない場合があるという課題も確認された。

5.3 δ -ナッシュ均衡戦略

ナッシュ均衡戦略を自然な着手に感じやすい理由として、相手に付け入れられない性質が挙げられる。しかし、当然ながら相手に付け入れられない戦略に対しては人間プレイヤは読みによるアドバンテージを得る戦略を考えることができないため、読み合いの演出に適しているとは言い難い。そこで、ナッシュ均衡戦略の性質をある程度維持しながら、強さを人間に近づける方法としてナッシュ均衡戦略を求める利得行列に、ノイズを付与する手法を提案する。人間プレイヤは、 5×5 通りの行動に対して、正確に利得を評価することは難しく、真の勝率が例えば 60% などの局面を人間は 50% などに間違えてしまうことは多々ある。このような人間の不正確さを導入した誤認知利得行列からナッシュ均衡戦略を求めることで、ナッシュ均衡戦略の性質をある程度維持しながら、強さを調整することができる。

本研究では、ナッシュプレイヤの強さや着手戦略を調整

する手法として、以下に示す δ -ナッシュ均衡戦略を提案する。 δ -ナッシュ均衡戦略は、ノイズを付与した誤認知利得行列からナッシュ均衡戦略を求める手法である。 δ -ナッシュ均衡戦略を求める手順について以下に示す。

- i 行 j 列の利得行列の要素 p_{ij} について、それぞれ一様分布 $(-\delta, \delta)$ などに従うノイズを付与する。ノイズを付与した利得行列から 5.2 節の手順 3 で用いた線形計画法による方法でナッシュ均衡戦略 N を計算する。
- 上記の試行を n 回行う。 k 試行目のナッシュ均衡戦略を N_k として、 δ -ナッシュ均衡戦略 $N' = \sum_k N_k / n$ を求める。

5.4 確率的な着手に関する評価

後退解析によって求めた勝率をそのまま用いたナッシュ均衡戦略は、人間ならば選択してもおかしくない行動をまったくとらないことがある。例えば、自分 2 行動、相手 1 行動に単純化した例で、真の勝率が 50% の手と 45% の手があった場合、後者は絶対に選択されない。しかし、人間にとっては真の勝率を正確に計算することは難しいことが多いため、実際には人間プレイヤが後者を選択してもおかしくない。

一方で δ -ナッシュプレイヤは、より人間に近い勝率評価を行うため、このような人間であれば有力に見える着手の確率を増加させることが期待できる。そこで、このような着手確率のバラつきについて検証を行った。

確率的な着手戦略を評価するため、簡易版ポケモンバトルの初期局面を用いる。簡易版ポケモンバトルは、プレイヤー間のパーティー入れ替えを加味しなければ、9 つの初期局面が存在する。これらの初期局面*3においてノイズ δ を変化させた場合の δ -ナッシュ均衡戦略を確認することで、着手戦略のバラつきを評価した。

$$\begin{bmatrix} 0.42 & 0.00 & 0.42 & 0.08 & 0.54 \\ 1.00 & 0.59 & 1.00 & 0.08 & 0.38 \\ 0.42 & 0.00 & 0.42 & 0.22 & 0.25 \\ 0.00 & 0.09 & 0.00 & 0.16 & 0.02 \\ 0.00 & 0.00 & 0.01 & 0.16 & 0.17 \end{bmatrix} \quad (1)$$

式 (1) はある初期局面における真の勝率による利得行列である。この利得行列から求めたナッシュプレイヤ、一様分布 $(-\delta, \delta)$ でノイズを付与した $n = 10$ の δ -ナッシュプレイヤの着手戦略の組 $(a_1, \dots, a_5), (b_1, \dots, b_5)$ を表 3 に示す。

表 3 のナッシュ均衡戦略は、 a, b の戦略ともに 2 つの行動のみ確率的に着手する戦略になっている。一方で δ -ナッシュ均衡の着手戦略は $\delta = 0.05$ でも 3 つ以上の行動を選択する可能性がある。人間的には選択してもおかしくない行

*2 実際には step3,4 ですべての状態のナッシュ均衡戦略は求まっているが、本研究ではコンピュータのメモリ (RAM) を節約するためこのような実装とした。

*3 この実験ではパーティー 2 の設定は表 2 のポケモン E、ポケモン F の HP を 5 にした設定を用いた。

表 3 初期局面における各プレイヤーの着手戦略

δ	着手戦略の組み合わせ
0	(0.00, 0.30, 0.70, 0.00, 0.00), (0.00, 0.19, 0.00, 0.81, 0.00)
0.05	(0.00, 0.27, 0.52, 0.11, 0.09), (0.00, 0.18, 0.00, 0.80, 0.02)
0.1	(0.01, 0.25, 0.39, 0.13, 0.22), (0.00, 0.18, 0.00, 0.76, 0.06)
0.2	(0.08, 0.43, 0.35, 0.08, 0.06), (0.00, 0.19, 0.00, 0.74, 0.07)

動は δ が低い場合には選択確率が低いものの、 δ が増加するにつれ、選択確率も増加している。また、一部の行動については δ が大きい場合でも着手確率が 0 になっており、典型的な悪手については選択しないことが確認できる。検証では例示した局面以外のすべての初期局面についても同様の傾向を確認している。

以上の検証結果から δ -ナッシュプレイヤーは、人間プレイヤーの誤認知を模倣することで、少し損する手を含んだ人間らしい多様な着手を行うことができると考える。

5.5 プレイヤーの強さと着手の自然さに関する評価

δ -ナッシュ均衡戦略に従う δ -ナッシュプレイヤーについて、ノイズのサイズを変化させることで勝率を調整できるかについて評価実験を行った。また、付与するノイズの性質などを変更させることで勝率や着手戦略にどのような影響があるかについても検証を行った。

以下に示す 2 種類のノイズを付与した $n = 10$ の δ -ナッシュプレイヤーについて、それぞれナッシュプレイヤーと 500 回対戦させ勝率を確認した。表 4 に対戦成績を示す^{*4}。

- 一様分布 $(-\delta, \delta)$
- 一様分布 $(\pm\delta(1 + 4p(1 - p)))$

表 4 ノイズの分布とパラメータごとの δ -ナッシュプレイヤーの勝率比較

ノイズ δ	0	0.05	0.1	0.2
ナッシュ ($\delta = 0$)	0.528	-	-	-
一様分布 $(-\delta, \delta)$	-	0.526	0.500	0.458
一様分布 $(\pm\delta(1 + 4p(1 - p)))$	-	0.516	0.460	0.378

5.5.1 ノイズサイズによる強さの調整

ナッシュプレイヤーとの対戦成績では、2 種類の分布ともノイズが大きくなるにつれ、勝率が下がっていることが確認できる。 δ -ナッシュプレイヤーは、ナッシュプレイヤーと比較して強さを抑えることに成功しており、プレイヤーの強さという観点では人間が自然に感じる着手の条件を満たしている。

5.5.2 ノイズ δ を変動させた際の比較

一様分布 $(\pm\delta(1 + 4p(1 - p)))$ は i 行 j 列の利得行列の要素 (勝率) p によって、ノイズのサイズが変動するモデル (以下、 δ 可変モデル) である。

δ 可変モデルは、人間プレイヤーの場合、真の勝率が極端

^{*4} 比較実験に用いている seed 値の影響で短期的にナッシュプレイヤーに 50%以上の勝率を示すことがある。

なときよりも、互角に近いほうが、真の勝率を見誤る可能性が高いことを反映させたモデルである。例えば、真の勝率が 0% のときに、勝率が 10% だと誤認することよりも、真の勝率が 50% のときに勝率 60% と誤認することの方が多いのは人間プレイヤーであれば自然といえる。

δ 可変モデルはノイズが大きい場合には、弱くなりすぎる問題があるため、ノイズが比較的小さい場合には、より人間らしい勝率を求める上で有効だと考える。

5.5.3 着手の自然さと有望な設定

強さと着手の自然さに関する評価から有望な設定について考察する。まず、より人間に近い誤認知を表現できる点で δ 可変モデルは有望なモデルといえる。強さについては、 $\delta = 0.05$ のモデルはナッシュプレイヤーとの差が小さいため、ノイズ δ は 0.1 から 0.2 が適切な範囲だと考える。

評価をまとめると、一様分布 $(\pm\delta(1 + 4p(1 - p)))$, $\delta = 0.1$ から 0.2 が読み合いを演出する前提となる標準着手として有望だと考える。ただし、人間プレイヤーの実力に応じて、 δ -ナッシュプレイヤーの強さを調整する必要性もあるため、有望な設定はこの限りではない。

6. バイアス着手生成機能

バイアス着手生成機能では、5.3 節で述べた δ -ナッシュ均衡戦略を求める際の利得行列の要素にバイアスを付与することを提案する。ノイズを付与した利得 p_δ に、さらにバイアスによる補正をかけた誤認知利得行列を作成する。この利得行列を用いて δ -ナッシュ均衡戦略を求めることで、人間のような癖や傾向を持った戦略を生成する。もとの勝率 p_δ に対して、式 (2) に示す関数の出力 p' を補正後の勝率とした。なお、式中における α はその行動の組み合わせを選好するパラメータで、負ならば避けようとする、正ならば好むことを表す。

$$p' = p_\delta + \alpha(1 + 20p_\delta(1 - p_\delta)) \quad (2)$$

この式では、 p_δ が 0 または 1 に近い場合は、 p_δ には小さい変異 (α 程度) しか与えられない。一方で、 p_δ が 0.5 に近い場合には、 p_δ には大きい変異 (6α 程度) が与えられる。はっきり負けや勝ちとなるような局面の評価はバイアスであり変わらないと考えるため、このような変異の差を与える式とした。

6.1 付与する 2 種類のバイアスとその結果

前節で提案した手法を用いて実際に人間プレイヤーのような傾向や癖を付与し、その結果を検証する。AI プレイヤーに実際に与えるバイアスを以下に示す。

攻撃/交換を好むバイアス

簡易版ポケモンバトルの行動は攻撃と交換に分類できる。人間プレイヤーの好みや性格、プレイスタイルによって攻撃/交換の選択比率が偏ることは十分に考え

られるため、これらの選択肢すべてにバイアスを付与する。具体的には、攻撃を好むバイアスの場合、行動 a_i が攻撃行動だった場合、 $p_{i1}, p_{i2}, \dots, p_{i5}$ に (2) 式で正の α を用いて利得の補正を行う。

有効/非有効な攻撃を好むバイアス

簡易版ポケモンバトルの攻撃行動はダメージの補正倍率 1.5 倍以上の攻撃（以下、有効攻撃）、補正倍率 1.0 倍の攻撃（以下、普通攻撃）、補正倍率 0.5 倍の攻撃（以下、非有効攻撃）に分類できる。このバイアスでは、相手プレイヤーが場に出しているポケモンに対して、有効/非有効な攻撃行動に対してバイアスを付与する。比較的素直なプレイヤーや、安定思考のプレイヤーは有効攻撃を好み、素直でないプレイヤーや先読みが得意なプレイヤーは非有効攻撃を好む傾向が想定される。

6.1.1 手法の有効性に関する評価

提案したバイアスが有効であるかについて、評価実験を行った。誤認知利得に、更にバイアスによる補正をかけた利得行列からナッシュ均衡戦略を求めるプレイヤーの 500 戦分の対戦ログを分析することで、手法の有効性を検証した。なお、検証時の各パラメータはノイズ $\delta = 0.05$ 、 $\alpha = 0.03$ とした*5。

図 2 に通常の δ -ナッシュプレイヤー、攻撃を好むバイアスを付与したプレイヤー、交換を好むバイアスを付与したプレイヤーの対戦ログの分析結果を示す。

通常の δ -ナッシュプレイヤーと攻撃を好むバイアスを付与したプレイヤーをそれぞれ比較すると、攻撃バイアスプレイヤーの攻撃選択確率が 8%以上増加していることが確認できる。また、交換を好むバイアスを付与したプレイヤーの着手分析結果と通常の δ -ナッシュプレイヤーを比較すると、交換選択確率が 15%以上増加している。

次に、攻撃行動のうち有効な攻撃を好むバイアスを与えたプレイヤー、非有効な攻撃を好むバイアスを与えたプレイヤーの着手分析結果を図 3 に示す。

通常の δ -ナッシュプレイヤーと、有効な攻撃を好むバイアスを付与したプレイヤーを比較すると、有効な攻撃を選択する確率は 26%以上と大幅に増加している。非有効な攻撃を好むバイアスを付与したプレイヤーについても、非有効な攻撃の選択確率が約 26%と大幅に増加している。

以上の結果から、定義したバイアスについて、それぞれ意図した行動の着手確率を増加させることができており、人間プレイヤーが持つような癖や傾向を表現することができたといえる。

さらに、人間のような癖や傾向を付与した δ -ナッシュプレイヤーの強さについても検証を行った。5.5 節の実験と同様に、ナッシュプレイヤーと 500 回対戦を行い、プレイヤーの勝率について確認した結果を表 5 に示す。なお、検証時の

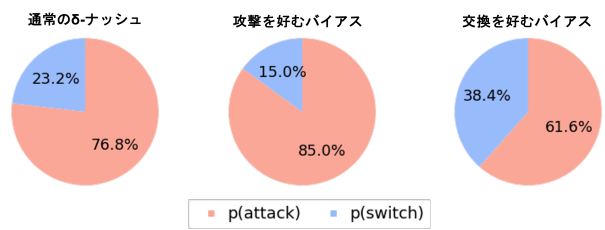


図 2 各プレイヤーの行動選択確率

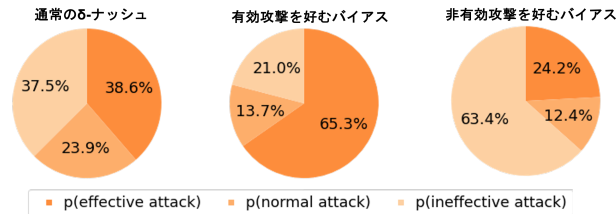


図 3 各プレイヤーの行動選択確率

表 5 各バイアスを付与した δ -ナッシュプレイヤーの勝率

δ -ナッシュ	攻撃	交換	有効攻撃	非有効攻撃
0.456	0.470	0.456	0.484	0.436

パラメータは $\delta = 0.05$ 、 $\alpha = 0.03$ とした*5。

各バイアスを付与したプレイヤーは、通常の δ -ナッシュプレイヤーと比較しても、勝率の差は比較的小さいことがわかった。このことから、人間的な癖や傾向を付与した δ -ナッシュプレイヤーは、 δ -ナッシュ均衡戦略の性質を一定程度保持しているといえる。

また、著者がテストプレイとして各バイアスを与えたプレイヤーと複数回対戦することで挙動を観察した。テストプレイでは、攻撃を好むバイアスを与えたプレイヤーであっても、例えば初期局面でタイプ相性が不利な場合は交換を行うなど、バイアスを与えた着手以外も状況に応じて用いることを確認した。

6.1.2 人間プレイヤーの着手との比較

人間プレイヤー（著者）の着手分析結果との比較を行う。図 4 に、ナッシュプレイヤー、人間プレイヤー（著者）、有効攻撃を好むバイアスを付与したプレイヤーの着手分析結果を示す。なお、図中の交換行動の選択確率は、場のポケモンに対して有利なタイプのポケモンに交換する確率と、それ以外のポケモンに交換する確率を示している。

人間プレイヤーの着手には、攻撃や交換で場のポケモンに対して有利になる行動を選択する傾向が見られる。しかし、ナッシュプレイヤーは、概ね均等に各攻撃行動や交換行動を選択しており、人間的な癖や傾向に近いとは言い難い。

一方で、有効な攻撃を好むバイアスを付与したプレイヤーは、人間の攻撃行動の選択確率に近い分布となっており、人間的な癖や傾向を表現できていることがわかる。各交換行動の選択確率に関しては、バイアスを与えていないためナッシュプレイヤーの分布に近い分布になっており、意図した結果になっているといえる。

*5 ノイズの分布には、正規分布 $(0, \delta(1 + 4p(1 - p)))$ を用いた。

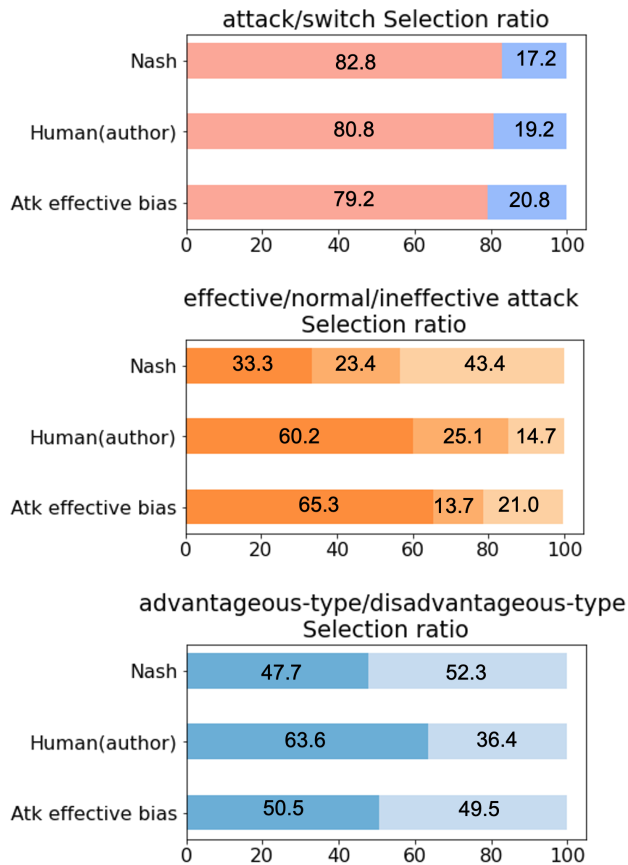


図 4 人間プレイヤーとバイアスプレイヤーの比較

7. 搾取戦略実施機能

本章では、人間のような読みを模倣した着手を生成する搾取戦略実施機能のアプローチについて述べる。

人間プレイヤーが行う行動の読みは大きく分けて、次の相手プレイヤーの行動予測と、予測される相手の行動に対してアドバンテージを得る戦略を考えることの2つから成る。本研究では前者を相手モデリングによる着手予測、後者については予測された行動に対して搾取戦略を用いることを提案する。

まず、相手モデリングによる着手予測に関するアプローチについて説明する。相手モデリングには、各バイアスを付与したプレイヤーによって多様な着手戦略を得られることを利用する。以下に相手モデリングによる着手予測の具体的な手順を示す。

- 現在または過去のゲームにおいて、各ターンごとに相手プレイヤーの着手と、攻撃/交換/有効攻撃/非有効攻撃の各バイアス、バイアス定数 α 、ノイズ δ をそれぞれパラメータとした δ -ナッシュ均衡戦略 (の組) を保持する。
- 相手の行動履歴 $a_i (i = 1, \dots, n)$ について、 k 番目の δ -ナッシュ均衡戦略 N_k の尤度 $L_k = \prod_i^n p(N_k | a_i)$ をそれぞれ求める。 $k^* = \arg \max_k L_k$ として最も尤度が

高くなるモデル N_{k^*} を求める。

- 現在の状態に対して、 N_{k^*} を用いて相手の行動確率分布を得る。

次に、予測される行動に対する搾取戦略の生成方法についても説明する。相手の戦略が固定されている場合、相手の戦略に対する最適応答反応は多くの場合、ある行動を確率 1 で選択する純粋戦略になることが知られている。そのため、真の勝率による利得行列をもとに、自身の着手の中で最も勝率が高い行動を確率 1 で着手することで、相手の行動に対してアドバンテージを得ることができると考える。

このアプローチによって人間プレイヤーのような「読み」を模倣する搾取戦略を強調することが可能だと考えたものの、評価実験によって有効性を示すまでには至らなかったため、搾取戦略実施機能の完全な実装とその評価については今後の展望とする。

8. おわりに

本研究では、ポケモンバトルを簡略化した同時手番ゲームを対象として、AI プレイヤーに人間のような読み合いを演出させる手法の確立を目指した。読み合いを演出する手法では、人間が自然に感じる標準着手として、 δ -ナッシュ均衡戦略を生成し、この戦略に人間のような癖や傾向をバイアスとして付与する手法の有効性を示した。また、人間の読みを模倣した着手についても相手モデリングと搾取戦略を組み合わせるアプローチを示した。

今後の展望として、提案した簡易版ポケモンバトル以外の同時手番ゲームや、隠れた情報がある不完全情報ゲームについても本研究の手法を応用し読み合い演出に適した多様な着手戦略を生成することが期待される。

謝辞 本研究は JSPS 科研費 JP20K12121 の助成を受けたものです。

参考文献

- [1] Silver, D. et al.: Mastering the game of Go with deep neural networks and tree search, *nature*, Vol. 529, No. 7587, pp. 484–489 (2016).
- [2] 池田心: 楽しませる囲碁・将棋プログラミング, オペレーションズ・リサーチ: 経営の科学 (2013).
- [3] 仲道隆史, 伊藤毅志: プレイヤーの技能に動的に合わせるシステムの提案と評価, 情報処理学会論文誌, Vol. 57, No. 11, pp. 2426–2435 (2016).
- [4] Van der Kleij, A.: Monte Carlo tree search and opponent modeling through player clustering in no-limit Texas hold’em poker, *University of Groningen, The Netherlands* (2010).
- [5] Sato, N., Tamsiririrkkul, S., Sone, S. and Ikeda, K.: Adaptive fighting game computer player by switching multiple rule-based controllers, *2015 ACT-CSI, IEEE*, pp. 52–59 (2015).