Vol.2018-ICS-191 No.6
2018/3/26

7並べにおける行動戦略

石川 諒人[1,a)]　藤田 桂英[1,b)]

概要：不完全情報ゲームにおいて，ポーカーや麻雀，大富豪などの研究が活発に行われている．それらに比べて 7 並べはまだ十分な研究が成されていない．本研究では，7 並べを題材に評価値法と UPP 法の拡張を考案し，手法を導入したエージェントによる対戦実験を行った．評価値法は局面から手に評価値をつける手法で，UPP 法は相手の行動から不完全情報を推測する手法である．実験の結果，提案した評価値法を使ったエージェントが既存手法に有意に勝ち越し，評価値法の有効性を確認した．また，拡張 UPP が既存手法であるモンテカルロ法に劣り，7 並べにおいて UPP 法による推測が有効でないことを示した．

キーワード：ゲーム，不完全情報，7 並べ

1. はじめに

人工知能の研究分野において，思考ゲームに関する研究は活発に行われている．これはゲームが現実世界をモデル化したものと捉えることが可能であり，現実と比較して人工知能の目的や問題設定が明確で評価が容易なためである．ゲームはプレイヤーに与えられる情報の観点から完全情報ゲームと不完全情報ゲームの二つに分けることができる．不完全情報ゲームは不完全情報を含むため強力な AI を作成することが完全情報ゲームに比べて一般的に難しいが，ゲームをより抽象化した現実における問題解決に応用できる可能性を秘めている．以上から不完全情報ゲームの研究は人工知能の研究テーマとしては大きな意義を持っている．

チェスや将棋，囲碁などの完全情報ゲームでは，すでに人間を越える AI が完成されている．例えば，将棋においては様々なイベントでトッププロ棋士に互角以上の成績を残し，トッププロ棋士を上回る実力をコンピュータ将棋があると分析され [1]，すでに情報処理学会からコンピュータ将棋プロジェクトの終了宣言 [2] が出されている．さらに，将棋よりも難しいとされる囲碁においても，AlphaGo がすでに人間のトッププレイヤーに勝利し，それを上回る実力を持つ AlphaGo Zero[3] も開発されている．加えて，AlphaGo Zero を汎化して他のゲームでも動作するプログラム Alpha Zero[4] がチェスや将棋において既存の強い AI に勝利した．

[1] 東京農工大学　工学部　情報工学科
[a)] ishikawa@katfuji.lab.tuat.ac.jp
[b)] katfuji@cc.tuat.ac.jp

Noam Brown らは不完全情報ゲームの一つであるノーリミットヘッズアップテキサスホールデムポーカーにおいて，ゲームの抽象化とサブゲームを用いた学習，自己修復によるブループリントの強化を行うことで，4 名の人間のプロプレイヤーたちに勝利した [5][6]．人工知能の課題の一つとして，どのように不完全情報ゲーム特有の課題を乗り越えていくかが注目されている．このように不完全情報ゲームにおいて，ポーカーや麻雀，大貧民などの一部ゲームの研究が活発に行われている．しかし，現実における問題は多種多様であり，一部のゲームの研究により問題すべてが解決がされるわけではない．今後，さらに研究が必要なゲームとして 7 並べがあげられる．7 並べは日本において一般的なゲームであるにも関わらず，目立った研究成果は発表されていない．しかし，多人数性や情報の不完全性，推測困難性，公開される情報量の少なさなどの特徴を持ち，ポーカーや麻雀とは異なるジャンルのゲームに分類される．そのため，7 並べは十分に研究に値するテーマであるといえる．

本研究では 7 並べを取り上げ，現状態から手に対して評価値を決める評価値法，過去のプレイアウトから非公開情報を推測する UPP 法を提案する．その後，各アルゴリズムを搭載したプレイヤーを作成し既存手法と対戦させることで，提案手法が有効であるかを確認する．

以下に本論文の構成を示す．まず，ゲーム研究の基礎となる理論および先行研究について示す．次に，本論文で扱う 7 並べのルールについて示す．その後，UPP アルゴリズムの 7 並べにおける拡張について示し，一般的な UPP アルゴリズムの概要について述べたのち，7 並べにおける

UPP アルゴリズムの適用例について示す．その後，ゲーム AI で有効な幾つかの手法を 7 並べに導入し，それを用いて行った実験結果について示し，本論文のまとめを示す．

2. 関連研究

ゲーム研究においてよく用いられるアルゴリズムについて述べる．

乱択法

文字どおり合法手の中から乱数を用いてランダムに手を選択する手法である [7]．合法手の数が少なければある程度の強さを誇るが，多くなると最善手をとる可能性が低くなってしまう．また，人間であれば一目で悪手であると判断できる手を選択してしまう場合もある．

ヒューリスティック

厳密な論理ではなく，経験や直感から暫定的もしくは近似的な判断を行うのことをヒューリスティックという．必ず正しい方法を導くことは出来ないが，ある程度の精度で良い方法を導くことができる．例えば，大富豪で手札の組の数が少なくなるように組を生成し，自分が次の番で出せる組が多くなる組もしくは提出することで場を流せる可能性が高い組を提出する戦略がヒューリスティックの例である [8]．

ルールベース法

比較的単純なルール (=規則) に基づいて判断をするアルゴリズムのことをルールベース法という．「もし〜なら · · · する」というルールを組み合わせることで実現される．

状態価値評価法

状態価値評価法は，場の状態を評価することで次の行動を決定する，もしくは決定に用いる要素を生成する．例えば，将棋においては駒に価値を設定して評価することがある [9]．

モンテカルロ法

ゲーム木の一部を乱数を用いてランダムにシミュレーションして手を選ぶことをモンテカルロ法という [7]．ゲームの特徴によらず適用できるためよく用いられる．終局まで乱数を用いてシミュレーションしプレイすることをプレイアウトといい，何度もプレイアウトを繰り返し報酬 (勝敗や得点) をもとに次の手を決定する．乱数を用いるため精度を出すためにはシミュレーション回数を増やす必要があるが，ラスベガス法 [7] とは異なり正しくない解が返されることも有る．完全ゲーム木をすべて探索できない多くのゲームで活用される．

また，不完全情報ゲームにおいては，情報が定まっていないのでそのままではシミュレーションをすることが出来ない．そのため，図 2 のように，情報を仮定して世界を決めてからシミュレーションを行うことでモンテカルロ法を実現する [10]．

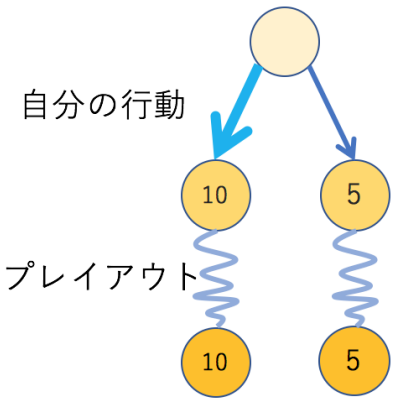


図 1　モンテカルロ法

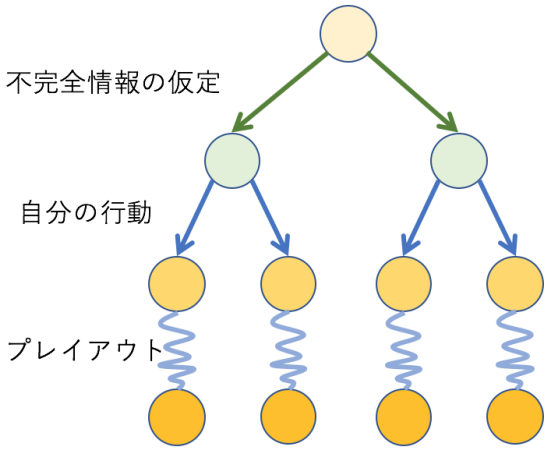
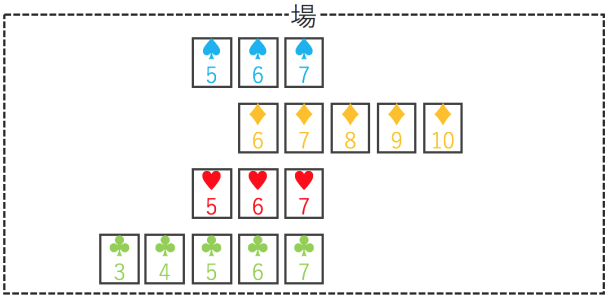


図 2　不完全情報ゲームにおけるモンテカルロ法

3. 7 並べ

7 並べは，トランプを使用したゲームの一種であり，英語では sevens などと呼ばれている．また，ファンタンという 7 並べの元となったゲームが存在する．多人数零和不確定不完全情報ゲームであり，多人数性や不確定性，不完全情報を含むことから難易度が高く他のゲームと比較して十分な研究が成されていない．しかし，これらの多人数性や不確定性，不完全情報性，推測困難性，公開される不完全情報量の少なさは現実問題に近いことから研究に値するテーマであると考える．

ゲームは図 3 のように，一般に 3 から 6 人程度で行われ，各プレイヤーは自分の手札と場に出されているカード，他プレイヤーのカード枚数のみを知ることができる．まず，プレイヤーは開始時に配られた手札からすべての 7 を場に出す．その後，プレイヤーは 7 と隣り合う数もしくは再帰



プレイヤー	A	B	C(自分)	D
手札枚数	8	10	8	10
残りパス回数	2	2	3	1

図 3 7 並べの局面例

的に「7 と隣り合う数」と隣り合う数のカードを出していき，先に手札がなくなったプレイヤーから順位が決まっていく．なお，カードを出さずにパスすることも可能である．

3.1 ルールと用語

7 並べにはいくつかローカルルールが存在する．そのため，本研究では一般的かつ簡単なルールを基準として研究を進める．その具体的なルールを次に示し，また，7 並べで利用される用語についても説明する．

基本的なルール

ゲームを行うのに必要なプレイヤーの人数は 4 人である．ゲームに使用するトランプのカードは 4 種類のスート (スペード・ダイヤ・ハート・クローバー) の 1 から 13 までの 52 枚である．席順はランダムに決定され，カードはランダムに選ばれた一番手から順に配布される．ゲーム開始前にすべてのプレイヤーの手札の 7 はすべて場に出され，すべてのプレイヤーは 3 回のパス回数が与えられる．

カードの出し方

場に出すカードは，同じスートで 7 と隣り合う数もしくは再帰的に「7 と隣り合う数」と隣り合う数のカード 1 枚でなければならない．なお，ローカルルールで，「あるスートの 7 から 1(もしくは 7 から 13) がすべて場に出されたときそのスートの 13 側 (1 側) から出せる」というものが存在するが本研究では準拠しない．

一番手から順に席順に従ってカードを出していく．プレイヤーは自分の手番となったときにカードを出すかパスをするか選択する必要がある．出せるカードがない場合には強制的にパスが選択される．パスをすると，次のプレイヤーに手番が渡り自分のパス回数が一つ減る． パス回数が 0 のときにパスをするとリタイアとなり，既にリタイアした人を除いた最下位の順位となる．リタイアしたプレイヤーの手札は隣り合ってるか否かに関係なくすべて場に出される．リタイアせずにすべてのカードを出し終えて手札が 0 枚になった人から上がりとなり，すでに上がりの人を除いた最上位の順位となる．

得点

一回のゲームの点数は，順位が最下位のプレイヤーに 0 点，3 位のプレイヤーに $\frac{2}{3}$ 点，2 位のプレイヤーに $\frac{4}{3}$ 点，1 位のプレイヤーに $2(=\frac{6}{3})$ 点を与える．なお，点数が $\frac{2}{3}$ 点刻みなのは，正規化して全プレイヤーの得点の平均を 1 にするためである．

系列

あるスートの 1 から 6，もしくは 8 から 13 のカードの集合，またはそれに属することを意味する．例えば，スペードの 3 はスペードの 1 から 6 の系列であると表現することができる．

止める

出せるカードを意図的に出さずに，他のカードを出したりパスをしたりすることを意味する．止めることで，他者の行動の選択肢を制限することができる．

4. 7 並べにおける戦略アルゴリズム

本節では 7 並べにおける戦略の異なるアルゴリズムについて述べる．なお，評価値法と拡張型 UPP 法が提案手法であり，人間が一般に用いたり他のゲームに用いられる乱択法とルールベース法，モンテカルロ法が既存手法である．

4.1 7 並べにおける乱択法

7 並べに乱択法 [7] を導入する．7 並べにおける乱択法では，まずパスを除いた可能な手を取り出す．その後，それらの手から乱数を用いてランダムに手を選択する．パスを除く理由は、含めた場合にパス回数が試合の序盤で使い切ってしまってリタイアする可能性が高くなってしまうことを防ぐためである．

図 3 を例とすると，パスを可能な手はスペードの 4，ダイヤの J，クローバーの 8 の 3 つであり，乱数を用いてランダムに選択するので可能な各手が出る確率は $\frac{1}{3}$ ずつである．

4.2 7 並べにおけるルールベース法

7 並べにおけるルールベース法について述べる．7 並べにおけるルールベース法では，パスを除いた可能な手の中から，7 と最も離れた数値のカードを取り出す．その後，その中から乱数を用いてランダムに選択する．一般に人間がよく用いる手法の一つである．

例として図 3 の状況では，パスを覗いた可能な手の中から 7 と最も離れた数値のカードはダイヤの J であり，そのダイヤの J を選択する．

4.3 7 並べにおける評価値法

評価値法は，合法手に対して評価を行い，その値が最も高い手を選択する．設計の意図としては，自分の出せる手を増やし他者の出せるカードを増やさないように，各可能な手について評価する．具体的な評価方法は以下のとおりである．

自分の持っているパス以外の合法手の評価の決定方法は式 4-1 に従う．ただし，手札が残り一枚でそのカードを出せるときは，評価値によらずにカードを出すことを選択する．

式 4-1 合法手の評価方法

$a_i = \exp(r \cdot c_{own} - c_{others})$
a_i : 合法手 i の評価値
c_{own} : 合法手 i を除く，i の系列の自分が持っているカード枚数
c_{others} : i の系列の他プレイヤーが持っているカード枚数
r : まだ手札の残っている他プレイヤーの人数

なお，自分のパスの残り回数が他者よりも多く，すべての系列を止めているもしくは出し尽くされている場合，パスに対する評価値を無限大にする．それ以外の場合で，パス回数が残っているときはパスに対する評価を 0.99 に，残っていない場合は 0 にする．

図 3 を例として説明する．可能な手の評価値はそれぞれ，スペードの 4 は $\exp(4 \cdot 1 - 2) = e^2$，ダイヤの J は $\exp(4 \cdot 0 - 2) = e^{-2}$，クローバーの 8 は $\exp(4 \cdot 1 - 4) = e^0 = 1$，パスは 0.99 である．したがって，評価値が最も高いスペードの 4 を選択する．

4.4 7 並べにおけるモンテカルロ法

7 並べにおけるモンテカルロ法では，ランダムに手札を配布してシミュレーションを行い，最も得点の期待値の高い手を選択する．このシミュレーションでは，すべてのプレイヤーの戦略に乱択法など他の手法を適用する．

図 3 を例として説明する．可能な手の得点の期待値がそれぞれ，スペードの 4 は 1.2，ダイヤの J は 0.9，クローバーの 8 は 1.5，パスは 0.7 である場合，期待値が最も高いクローバーの 8 を選択する．

4.5 7 並べにおける UPP アルゴリズム

本項では 7 並べにおける UPP アルゴリズムについて述べる．不完全情報ゲームにおいて世界を仮定する際に，仮定された世界が実際の世界と異なる可能性があること，仮定した世界の数だけ探索対象の数も増えることから良い手を選択することが困難である．この問題を緩和するために，本アルゴリズムでは過去の着手とプレイアウト結果からより起こりうる可能性の高い世界を見つけ出すことを目的とする [11].

4.5.1 既存の UPP アルゴリズム

まず，UPP アルゴリズムがどのように動き，世界の仮定や着手を決定しているかを簡単に述べる．既存の UPP では，二人ゲームを前提としており，入力として現在の局面の状態と一手前と二手前の着手，二手前の着手に用いたプレイアウト結果を受け取り，出力として世界の評価値を返す．大まかに分けると 5 つのステップから構成される．

ステップ 1:着手確認ステップ

現在の手番の直前の相手の着手 (一手番前) と直前の自分の着手 (二手番前) における着手を確認する．

ステップ 2:プレイアウト結果の参照・比較

直前の自分の手番で着手決定に用いたプレイアウト結果を確認する．それらのプレイアウトの世界それぞれにおいて，最初の自分の着手と次の相手の着手が実際のものと等しいプレイアウト結果を参照する．その後，より相手の勝率 (=得点の期待値) が高くなる世界を可能性の高い世界とするように更新値を設定する．

ステップ 3:不完全情報の推定ステップ

各世界の予め設定された評価値を更新する．ステップ 2 で得られた更新値に応じ，世界の評価値に一定の値を加算する．そのため，世界の評価値は可能性が高いほど高くなる．

ステップ 4:重み付きモンテカルロ法

ステップ 3 で更新された世界の評価値を利用し，重み付きモンテカルロ法を行うことで着手を決定する．世界の評価値と割り振られるプレイアウトの数は単調増加の関係とする．なお，このプレイアウト結果は保存し，次回の着手決定に利用する．

ステップ 5:世界の刈り取りステップ

決定した着手により不完全情報の一部が明らかになる場合がある．その情報が矛盾する世界にプレイアウトを割り振らないようにする．

4.5.2 UPP の 7 並べへの拡張

既存の UPP アルゴリズムにおいて仮定しているゲームは，プレイヤーが二人であり世界の数が少なく探索可能なサイズである．一方，本研究で扱う 7 並べは多人数であり世界の数は最大で ${}_{39}\mathrm{C}_{13} \cdot {}_{26}\mathrm{C}_{13} \approx 8.4 \cdot 10^{16}$ と莫大で，すべてを探索することは不可能である．したがって，7 並べにおいて，UPP アルゴリズムを適用させるために，UPP アルゴリズムを拡張する．

まず，原始 UPP では不完全情報の組み合わせを世界としていた．一方，7 並べ拡張版 UPP では，カードごとに誰に配られたかを世界とし，それぞれの世界に評価値を設定する．

ステップ 1:着手確認ステップ

現在の手番のすべての相手プレイヤーの直近の着手と自分の直前の着手を確認する．

ステップ 2:プレイアウト結果の参照・比較

前の自分の手番で着手決定に用いたプレイアウト結果を確認する．それらのプレイアウトの各世界において，最初の自分の着手とすべての相手プレイヤーの次の着手が実際のものと等しいプレイアウト結果を参照する．その後，より相手の勝率が高くなるカードを可能性の高い世界とするように更新値を決定する．

ステップ 3:不完全情報の推定ステップ

前項で得られた比較結果から世界の評価を行う．世界の評価値は式 4-2 によって定める．世界はそれぞれ最初に評価値 1 を持っており，もし世界が存在しているなら，世界の評価値は初期値にその世界がありそうだと評価された分だけ値を加算することで求めることができる．c は定数であり，この値が大きいほど一回の評価で評価値が大きく変化する．

式 4-2 世界の評価値

$$W_n = \begin{cases} 1 + t_n c & (F_n = 1) \\ 0 & (F_n = 0) \end{cases}$$

n : カードの番号
W_n : カード n の評価値
t_n : カード n の持つ値
c : 加算値
F_n : 世界 n が存在するか (1:存在する 0:存在しない)

t_n は式 4-3 のようにして求めることができる．これは評価値を求めるのに必要な値である．

式 4-3 t_n を求める式

$t_n = t_n +$ (世界 n におけるプレイアウトの得点) $- 1$
t_n : 世界 n の持つ値．初期値 0

ステップ 4:重み付きモンテカルロ法

前項で求めた世界の評価値を利用してモンテカルロ法を行う．モンテカルロ法では，まず式 4-4 の確率でカードを一枚ずつ配ることで，未知のすべてのカードを配りプレイアウトするすべての世界を決定する．

式 4-4 プレイアウトの可能性の式

$P(W_n) = \frac{W_n}{\Sigma_k W_k}$
W_n : 世界 n の評価値
$P(W_n)$: あるカードがある人に配られる確率

プレイアウトを割り振る世界が決定したら，可能な手の中からランダムに一つ選ぶ．可能な手に対してプレイアウトを行い，再び式 4-4 の確率に基づきプレイアウトする世界を決定する．決められた回数になるまでこれを繰り返し，最後に各可能手の各世界における点数を合計し，各可能手の合計プレイアウト回数で割ることで勝率を求める．最終的に，勝率が最も高かった手を着手として選択する．なお，次の手を決定する際のレイアウト結果の参照・比較ステップで利用するため，これらのプレイアウト結果と過去の手番を保存しておく．

表 1　既存手法と評価値法の対戦結果の得点の平均

プレイヤー		得点
乱択型	**ルール型**	0.93 : 1.07
ランダム型	**評価値型**	0.79 : 1.21
ルール型	**評価値型**	0.82 : 1.18
乱択型	**乱択 MC**	0.75 : 1.25
ルール型	**ルール MC**	0.79 : 1.21
評価値型	乱択 MC	1.08 : 0.92
評価値型	ルール MC	1.06 : 0.94
評価値型	評価値 MC	1.07 : 0.93

ステップ 5:世界の刈り取りステップ

このステップでは，過去の手番により明らかとなった情報により存在し得ない世界であると判断できる場合は世界の F_n 値を 0 にする．このステップにより，存在し得ない世界に対してプレイアウトが割り振られることがなくなる．このようにしてゲーム開始時に莫大な組み合わせ存在していた世界は手番が進むごとに刈り取られていき，より効率よくプレイアウトの割当を行うことができる．

例えば，すでにハートの 5 が場に出ているもしくは自分が持っている場合は，ハートの 5 が他者の手札に割り振られた世界の F_n 値を 0 にする．

5. 複数の 7 並べプレイヤープログラムにおける評価実験と考察

評価実験のために，4 節で述べた戦略の異なる複数のプレイヤープログラムを作成した．

5.1 複数の 7 並べプレイヤープログラムにおける評価実験

作成した AI プレイヤーの強さの程度の確認を行うために，プレイヤー同士で対戦させた．なお，実験結果に関して，プレイヤーの得点の優劣は人数の比によらなかったため，本項では 2 対 2 のみものを示し，別の人数の組み合わせは付録に記載する．また，表内ではルールベースをルール，モンテカルロを MC と略して称する．さらに，以降の対戦結果はすべて，ウェルチの t 検定より対戦したプレイヤー同士の得点の平均値は有意水準 1%で差があることを確認した．

5.1.1 既存手法と評価値型の対戦結果

乱択法とルールベース法，評価値法に加えて，それらにモンテカルロ法を適用したプレイヤーを組み合わせて対戦を行った．それぞれの組み合わせで 1000 試合行い，得点の平均を表 1 に示す．なお，一つの手番におけるモンテカルロ法の思考時間は 1 秒とした．

5.1.2 モンテカルロと UPP の対戦結果

乱択型とルールベース型のモンテカルロと UPP の AI 同士で対戦を行った．実験の条件は同様であり，その結果を表 2 に示す．なお，試合回数 1000 回で思考時間 1 秒とした．

また，UPP アルゴリズムのパラメータ c を $c = kx$ として，k を 1 から 10 の間を 0.5 刻みで動かしながら思考時間 1 秒で試合回数 100 回で予備実験を行い，最も得点平均が高かったときの値である 5 を c の値として採用し実験を行った．なお，直前のシミュレーション回数によらないように，x は直前のシミュレーション回数の逆数とした．

表 2 モンテカルロと UPP の対戦結果の得点の平均

プレイヤー		得点
乱択 MC	乱択 UPP	1.14 : 0.86
ルール MC	ルール UPP	1.20 : 0.80

5.2 複数の 7 並べプレイヤープログラムにおける考察

5.1 で行ったプレイヤーの対戦実験の結果を考察する．

5.2.1 既存手法と評価値型の対戦の考察

表 1 より，評価値型プレイヤーが乱択型やルールベース型よりも強いこと，乱択法とルールベース法はモンテカルロ法を適用した方が強いこと，評価値法はモンテカルロ法適用しない方が強いこと，評価値法が最も強いことが分かる．

まず，評価値法がルールベース型よりも強い理由について考察する．ルールベース型プレイヤーは，そのアルゴリズムから出せるカードのみについて評価している．一方，評価値型プレイヤーは出せるカードのみならず，手札や場のカードについても考慮して評価している．また，設計で述べたとおり，自分の出せる手を増やし他者の出せるカードを増やさないように各可能な手を評価するという意図している．これらの評価対象の範囲の差と設計が戦略に大きな違いを生み，乱択型は無論のこと，ルールベース型プレイヤーとの強弱の差ができたと考えられる．

次に，乱択法とルールベース法に関してモンテカルロ法を適用した方が強い理由について考察する．モンテカルロ法では，乱数を用いてランダムに手札を配りシミュレーションを行い，次の各手が平均的に良いか否かを評価することができる．したがって，ランダムやランダムに近い戦略を取る場合，モンテカルロ法を適用した方が強いプレイヤーになると考えられる．

次に，評価値型においてモンテカルロ法を適用しない方が強い理由について考察する．4.3 で述べたとおり，評価値型プレイヤーは場全体を考慮し評価することで手を選択しているので，試合途中の評価値型プレイヤーの手札はランダムに配った場合と比較すると良い状況であると予想される．一方，モンテカルロ法では世界の仮定の際に他者のプレイヤーの手札をランダムに配ることで実現している．その結果，評価値型モンテカルロプレイヤーは他プレイヤーの手札を実際と比べて悪い状態で仮定しシミュレーションを行うために，誤った評価を行ってしまうと考えられる．したがって，不完全情報ゲーム全般に共通することでは有るが，7 並べにおいてモンテカルロ法を適用する場合は特に不完全情報を推測することが重要である．

最後に，評価値法が最も強いことについて述べる．モンテカルロ法は他のゲームによく用いられてベースラインとされることが多いが，本研究で提案した評価値法はそれを上回り優れたアルゴリズムであるといえる．また，評価値型がすべてのモンテカルロ型よりも強い理由については，評価値型が単に強い戦法であることに加えて，先にで述べたのと同様にシミュレーション時に仮定する手札に対して評価値型の手札は良い状況なため誤った評価を行うからだと考えられる．

5.2.2 モンテカルロと UPP の対戦の考察

5.1.2 より，乱択型とルールベース型は UPP 法を適用するよりもモンテカルロ方を適用する方が強いことが分かる．このことから UPP を 7 並べに拡張しても有効でなく，ガイスターで推測が成功したことから 7 並べにおける不完全情報の推測は困難である．7 並べにおける推測の難しさは次の二つが要因として考えられる．

- 推測の手がかりに比べて不完全情報の量が多い
- 他者の行動が他の手と比較してどれだけ良いか不明

一つ目の推測の手がかりに比べて不完全情報の量が多いことについて説明する．7 並べはゲームの性質により不完全情報の量が多く，最大 39 枚のカードが存在し相手の手札の組み合わせは 10^{53} 通りである．それに比べて 1 ターンに公開される情報の量であるカードは少ない．この手がかりと推測対象の差が 7 並べにおける推測の難しさの一要素である．

二つ目として，他者の行動が他の手と比較してどれだけ良いか不明な点について述べる．あるプレイヤー視点で自身の手札は既知の情報である．対して，他者視点では不完全情報である．また，7 並べというゲームのルール上，場に出せるカードの中で手札から合法的に場に出せるカードは一部である．これらから，あるプレイヤーの行動が他の手に比べてどの程度いい手か分からず，選択した理由を推測することが困難である．例えば，手札の状況が非常によくて意図的にパスを選択したのか，それとも出せるカードがなくやむを得ずパスをしたのかをその手のみを判断材料に他者視点から判別することは不可能である．

6. まとめ

本論文では，7 並べを題材に評価値法と UPP 法の拡張を提案した．既存手法を含めて思考時間 1 秒で 1000 回の

試合の対戦実験を行った．その結果から，既存手法よりも提案した評価値法が強いこと，UPP 法による推測が有効に働かないことが明らかとなった．また，7 並べにおいては推測が重要であるが，推測の手がかりに比べて不完全情報の量が多く，他者の行動が他の手と比較してどれだけ良いか不明であるために推測が困難であることも分かった．

今回の実験で用いた AI は，比較的シンプルなアルゴリズムが多いが，現在ではニューラルネットワークなど高度な機械学習や強化学習手法を用いているものが多い．特に，Alpha Zero や Libratus のようにニューラルネットワークを活用した AI の研究が盛んに行われ結果が出ているので，7 並べに対しても同様にニューラルネットワークを導入することが検討される．

また，7 並べにおける未知のカードの配り方は非常に多く，ゲーム木の大きさも考慮すると現在の計算機では全探索は困難である．したがって，本稿で用いた UPP アルゴリズムと同様に相手の手札などを推測するアルゴリズムの研究が重要である．

参考文献

[1] 小谷善行：第 3 回将棋電王戦を振り返って : 3．コンピュータ将棋の棋力の客観的分析-人間のトップに到達したか？-，情報処理， Vol. 55, No. 8, pp. 851–852 (2014).
[2] 一般社団法人情報処理学会：コンピュータ将棋プロジェクトの終了宣言，¥url-http://www.ipsj.or.jp/50anv/shogi/20151011.html (2015).
[3] Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T. and Hassabis, D.: Mastering the game of Go without human knowledge, Vol. 550, pp. 354–359 (2017).
[4] Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K. and Hassabis, D.: Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm, *ArXiv e-prints* (2017).
[5] Noam Brown, T. S.: Libratus: The Superhuman AI for No-Limit Poker, *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pp. 5226–5228 (online), DOI: 10.24963/ijcai.2017/772 (2017).
[6] Brown, N. and Sandholm, T.: Safe and Nested Subgame Solving for Imperfect-Information Games, *Advances in Neural Information Processing Systems 30* (Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S. and Garnett, R., eds.), Curran Associates, Inc., pp. 689–699 (online), available from ⟨http://papers.nips.cc/paper/6671-safe-and-nested-subgame-solving-for-imperfect-information-games.pdf⟩ (2017).
[7] Motwani, R. and Raghavan, P.: *Randomized Algorithms*, Cambridge University Press, New York, NY, USA (1995).
[8] 田頭幸三，但馬康宏：コンピュータ大貧民におけるヒューリスティック戦略の実装と効果，情報処理学会論文誌，Vol. 57, No. 11, pp. 2403–2413（オンライン），入手先 ⟨https://ci.nii.ac.jp/naid/170000131096/⟩ (2016).
[9] 谷川浩司：谷川浩司の本筋を見極める，日本放送出版協会 (2007).
[10] Frank, I. and Basin, D.: *Optimal Play against Best Defence: Complexity and Heuristics*, pp. 50–73, Springer Berlin Heidelberg (1999).
[11] 三塩武徳：ゲームの不完全情報推定アルゴリズム UPP とそのガイスターへの応用，東京農工大学, 卒業論文 (2013).

付　　録

表 A·1　既存手法と評価値法の対戦結果の得点の平均

人数比		1：3	2：2	3：1
乱択型	**ルール型**	0.94：1.02	0.93：1.07	0.97：1.08
ランダム型	**評価値型**	0.83：1.06	0.79：1.21	0.78：1.67
ルール型	**評価値型**	0.84：1.05	0.82：1.18	0.81：1.57
乱択型	**乱択 MC**	0.72：1.09	0.75：1.25	0.81：1.58
ルール型	**ルール MC**	0.79：1.07	0.79：1.21	0.83：1.51
評価値型	乱択 MC	1.18：0.94	1.08：0.92	1.03：0.91
評価値型	ルール MC	1.09：0.97	1.06：0.94	1.03：0.92
評価値型	評価値 MC	1.03：0.90	1.07：0.93	1.11：0.96

表 A·2　モンテカルロと UPP の対戦結果の得点の平均

人数比		1：3	2：2	3：1
乱択 MC	乱択 UPP	1.10：0.96	1.14：0.86	1.08：0.77
ルール MC	ルール UPP	1.18：0.94	1.20：0.80	1.07：0.80