

タブーサーチを用いた麻雀における最適行動の探索

吉村 健志^{1,a)} 宝珍 輝尚^{2,b)} 野宮 浩揮^{2,c)}

概要 :

近年、将棋やオセロといった2人零和確定完全情報ゲームにおいて、トップ層の人間と同等の実力を持つプレイヤーが実現されている。しかし、ポーカーや麻雀といった不確定不完全情報ゲームでは、そのようなプレイヤーを実現することは難しい。本論文は4人零和不確定不完全情報ゲームである麻雀を取り上げる。そこで、本論文では、トップ層の人間に勝る麻雀プレイヤーを実現するために、多人数不完全情報ゲームにおける最適行動の探索を検討する。ここでは、トップ層の人間の牌譜の知識を用いず、シミュレーションによって最適解の探索を行うタブーサーチを用いる手法を提案する。タブーサーチを適用するのは、打牌選択の局面と鳴きの局面である。提案手法を実現し、一致率実験とベンチマークプレイヤーとの対局実験を行った。一致率実験によって、序盤の局面に対して、有効な和了形を最大85%で見つけることができ、さらに、対局実験によって、提案手法がベンチマークプレイヤーよりも優れているという結果が得られた。これらの結果から、タブーサーチによる最適解の探索が有効である可能性を示した。

Searching Optimal Movements in Mahjong Using Tabu Search

KENSHI YOSHIMURA^{1,a)} TERUHISA HOCHIN^{2,b)} HIROKI NOMIYA^{2,c)}

Abstract:

This paper proposes a search method of optimal movements in multi-player games with imperfect information in order to implement a Mahjong player exceeding human top players. The proposed method uses tabu search. It does not use the record of a game. Tabu search is applied to the states of discarding tiles and using the tile that other players discarded. The experiments were carried out in order to evaluate the proposed method. In the experiment of evaluating the rate of concordance of the usage of tiles, the maximum rate of concordance reached to 85%. This means effective winning hands could be found in the initial states. In the experiment of playing a game with benchmark players, it is shown that the proposed method is better than benchmark players. From these results, the possibility of the effective search of optimal solution by using tabu search was indicated.

1. はじめに

近年、将棋やオセロといった2人零和確定完全情報ゲームにおいて、トップ層の人間と同等の実力を持つプレイヤーが実現されている。しかし、ポーカー [1] や麻雀 [2][3][4][5][6][7]

といった不確定不完全情報ゲームで実現することは難しい。理由として、相手の手札の情報が公開されていないことや、確率的要素が存在することで状態数が膨大になり、可能性のある局面全体の探索が行えないことが挙げられる。今回、研究の題材として取り上げた麻雀は、4人零和不確定不完全情報ゲームである。麻雀は、上述の不確定不完全情報ゲームであることに加え、多人数ゲームである。多人数ゲームは、3人以上のプレイヤーが存在するゲームを指すが、このようなゲームは2人ゲームと違い、自分の利益を最大にする行動が最適な行動になるとは限らない。これらの性質から、麻雀における最適行動を決定することは

¹ 京都工芸繊維大学大学院 情報工学専攻
Graduate School of Information Science, Kyoto Institute of Technology

² 京都工芸繊維大学 情報工学・人間科学系
Information and Human Sciences, Kyoto Institute of Technology

a) m6622049@edu.kit.ac.jp

b) hochin@kit.ac.jp

c) nomiya@kit.ac.jp

非常に困難である。

そこで、本論文では、トップ層の人間に勝る麻雀プレイヤーを実現するために、多人数不完全情報ゲームにおける最適行動の探索について検討する。ここでは、トップ層の人間の牌譜の知識を用いずに最適な行動を決定するために、タブーサーチ [8] を使用する手法を提案する。知識を用いずに最適な行動を決定できるようになれば、研究が進んでいない他の不完全情報ゲームへの応用や、最終的には実世界への応用も期待できると考えられる。

2. 基本事項

2.1 麻雀

麻雀は、2人や3人のプレイヤーで行うこともあるが、一般的には4人のプレイヤーで行う4人零和不確定不完全情報ゲームである。

各プレイヤーは、最初に与えられた手牌に対して、1枚牌を引くツモと呼ぶ動作、1枚牌を捨てる打牌と呼ぶ動作と相手の打牌を利用する鳴きと呼ぶ動作を順番に行い、点数を得られる形を構成し、和了することで点数を獲得する。和了の方法として、ツモを行って和了することもツモと呼び、他プレイヤーの打牌で和了することをロンと呼ぶ。

牌は34種類あり、各4枚ずつ存在するため、合計136牌存在する。牌には1から9の数字が記載されている数牌と一文字の漢字が記載されている字牌が存在する。数牌は3種類存在し、優劣は存在しない。字牌は7種類存在し、東、南、西、北の4種類を風牌、白、発、中の3種類を三元牌と呼ぶ。

手牌は通常14枚で構成され、和了するためには、3枚の牌で構成される面子が4組と2枚の同一牌で構成される対子が一組必要である。面子には3枚の同一牌で構成される刻子、4枚の同一牌で構成される槓子と連続する3枚の数牌で構成される順子がある。

以下に、本論文で使用する麻雀用語の説明を記載する。

聴牌 和了に必要な牌の枚数が1枚の状態。

向聴数 必要な牌が何枚引ければ聴牌になるかを表す数。

降り 自分の和了を諦め、相手の和了牌を打牌しないようにすること。

鳴き 相手の打牌と自分の牌を組み合わせて面子を構成すること。

有効牌 手牌に加えた時に、向聴数を小さくする牌。有効牌の総和を受け入れ枚数と呼ぶ。

流局 和了以外の要因で局が終了すること。流局時に聴牌していないプレイヤーは聴牌しているプレイヤーに罰符として点数を支払う必要がある。

染め手 1種類の数牌、または1種類の数牌と字牌のみで構成される和了形。

立直 鳴きを行わず、聴牌したときに、1000点を払うことで宣言することができる行為である。宣言後、手牌を

変化させることができなくなる。

役牌 3枚以上集めると役になる牌。三元牌と一部の風牌を指す。

喰いタン 鳴きを行って、タンヤオの役であること。タンヤオは、数牌の1、9牌と字牌を1枚も使わない役である。

ドラ 所持枚数分、役が付く牌であり、局の最初に決定する。役がドラしかない場合、和了することはできない。

赤ドラという牌も存在し、常にドラとして扱われる。

親 最初に打牌を行うプレイヤー。

面子候補 牌をあと1枚獲得することで面子を構成できる牌の集合。

孤立牌 面子、対子または面子候補にもならない牌。

半荘 各プレイヤーが親を2回ずつ行う対戦形式。1回ずつ行うことを東風戦と呼ぶ。

飛び 点数が0点未満になること。

ダブルロン 1つの捨て牌で2人のプレイヤーが和了すること。

連荘 親が和了、または、流局時に聴牌していた場合、もう一度親を行うこと。

2.2 タブーサーチ

2.2.1 概要

タブーサーチ [8] は組み合わせ最適化問題を解くときに使われるメタヒューリスティックな手法であり、1989年にGloverによって考案された。組み合わせ最適化問題には、巡回セールスマン問題、グラフ彩色、スケジューリング問題などが存在する。

このアルゴリズムでは、ある状態から複数の近傍値を探索し、最も良い近傍値に遷移する。この時、タブーリストと呼ぶ集合にこの遷移を記録しておき、この集合に存在する遷移は行わないことによって、ループを防ぎながら探索を行える。タブーリストに存在する遷移は行えないが、それ以外の遷移はどれだけ状態が悪くても行われるため、局所解への収束を防ぐことができる。

2.2.2 表記法

組み合わせ最適化問題を式 (1) で定式化する。

$$\text{Minimize } c(x) : x \in X \quad (1)$$

ここで、 $c(x)$ は線形、または非線形の目的関数である。 x は離散値であり、 X は x の上位集合である。

この問題を解くために、ある解 x から別の解を生成する遷移 s を X の部分集合 $X(s)$ を用いて式 (2) で定義する。

$$s : X(s) \rightarrow X \quad (2)$$

ある x に対して $s \in S$ を用いて、集合 $S(x)$ は式 (3) で定義される。

$$S(x) = \{s \in S : x \in X(s)\} \quad (3)$$

同様に, $X(s)$ も式 (4) で表される.

$$X(s) = \{x \in X : s \in S(x)\} \quad (4)$$

ここで集合 $S(x)$ は近傍関数とされる.

2.2.3 アルゴリズム

タブーサーチのアルゴリズムは, 以下の 4 ステップで構成される.

- (1) 初期値 $x \in X$ を選び, $x^* := x$ とする. 反復回数を $k = 0$ とし, タブーリスト T を空集合とする.
- (2) $S(x) - T$ が空集合であれば, ステップ 4 に進む. そうでなければ, $k := k + 1$ とし, $s_k(x) = \text{OPTIMUM}(s(x) : s \in S(x) - T)$ であるような最適な遷移 $s_k \in S(x) - T$ を選ぶ. $s(x)$ はある x が s によって遷移した値である.
- (3) $x := s_k(x)$ とする. $c(x) < c(x^*)$ であれば, $x^* := x$ とする. x^* は現在見つかった最適解である.
- (4) 反復回数が規定回数に達する, または, ステップ 2 からこのステップに遷移し, $S(x) - T$ が空集合であれば停止する. もしそうでなければ, タブーリスト T を更新し, ステップ 2 に戻る.

ここで, **OPTIMUM** とタブーリスト T を定義する必要がある. 通常, **OPTIMUM** は $s_k(x)$ を使って, 式 (5) で表される.

$$c(s_k(x)) = \text{Minimum}(c(s(x)) : s \in S(x) - T) \quad (5)$$

T の更新は式 (6) を用いて行う.

$$T := T - s_{k-t}^{-1} + s_k^{-1} \quad (6)$$

s^{-1} は遷移 s の逆であり, $s^{-1}(s(x)) = x$ を満たす. t はタブーリストの保持数である. $k \leq t$ の時, s_{k-t}^{-1} は存在しない.

式 (6) から, T は式 (7) で表される.

$$T = \{s^{-1} : s = s_h \text{ for } h \geq k - t\} \quad (7)$$

タブーリストの保持数は, 実験的に 5 から 12 の値が良いとされており, 特に 7 とする場合が多い.

3. 関連研究

3.1 知識を使用する手法

トップ層の人間の牌譜の知識を用いた研究について述べる.

水上らは, 平均化パーセプトロンによって, 1 人麻雀プ

レイヤを作成し, そのプレイヤに鳴きと降りの機能を追加することで, 平均的な実力の人間を上回るプレイヤを実現している [2]. また, 対戦相手のモデルを聴牌, 待ち牌, 得点の組み合わせとし, 個別に学習することで予測モデルを作成している. このモデルと 1 人麻雀プレイヤ [2], モンテカルロ法を組み合わせることで中級者程度のプレイヤを実現している [3]. しかし, このプレイヤは現在の順位と残り局数に応じた押し引きができないことや最終局に和了しても最下位が確定する手を作成するという問題点が見られた. これを改善するために, 牌譜中に現れた点数状況から最終順位を予測するモデルの学習を行い, モンテカルロ法のシミュレーションでの報酬に用いることで最終順位に基づく行動を行うようにしている. この結果, 中級者の実力を超えるプレイヤが実現されている [4].

三木ら [5] は, 牌譜をもとにツモの局面に対して, 実際に打たれた手とそれ以外の手の木構造データを抽出し, 木カーネルを用いた非線形 SVM (Support Vector Machine) によって手牌の分類を行っている. これを行った結果, 熟練者の打ち手との一致率は 51% となり, それなりに良い精度で最善手を予測できる分類器を得られていたため, 簡単な木構造をもとに木カーネルを利用するだけである程度分類が可能であることがわかった. しかし, 順子, 対子などの単純な構造のみで木を構築しているため, 重要な牌を見逃すという問題もあった.

3.2 知識を使用しない手法

トップ層の人間の牌譜の知識を用いない研究について述べる.

三木ら [6] は, モンテカルロ木探索を用いた麻雀プレイヤを提案している. 麻雀のゲーム木の分岐数は非常に膨大であるため, 全ての分岐を探索するのは不可能に近い. そのため, 相手の手牌や行動をランダムでシミュレートするために, モンテカルロ木による探索を使用している. この手法は, トップ層の人間の牌譜の知識を用いないにも関わらず, 向聴数を小さくする牌の集合からランダムで打牌を行うというプレイヤよりも成績が上回るという結果になっている. しかし, 鳴きの局面に対して, ランダムプレイアウトでほとんど和了することができなかつたため, 有効な報酬の大半が罰符となってしまっていた. そのため, 和了よりも聴牌を目指すような探索となり, 向聴数を減らせる鳴きが選択されやすくなる傾向があった.

高橋ら [7] は, 遺伝的アルゴリズムを用いて, 初期値と類似している和了形の探索を行った. 本手法は, ツモを行った局面を初期値として, 打牌とツモのシミュレーションを行っており, このシミュレーションによって得られた個体集団に対して, 初期値との類似度が最も高い個体を最終的な解としている. 上記操作によって得られた解を初期値として複数回適応させる広さ優先の探索と得られた解を次々

と初期値として適応させていく深さ優先の探索を行っており、深く広くすることで得られる解の類似度は高くなっているが、深さ 10 程度で計算に 5 分程度かかっており使用することは難しいという結論に至っていた。さらに、モンテカルロ法を用いた手法も提案している [7]。この手法は、初期値に対して、ランダムな打牌とツモのシミュレーションを規定回数行うことで和了形を探索するというものである。評価値として向聴数のみを用いたものと向聴数と受け入れ枚数を用いたものを比較した結果、後者の方が正解の打牌を選択する確率が高く、後者の有用性を示していたが、上級者に匹敵するプレイヤーは実現できなかった。

4. 提案手法

4.1 提案手法の概要

本研究では、多人数不完全情報ゲームである麻雀に対して、タブーサーチを適用する。適用した局面は、打牌選択と鳴き選択の局面である。タブーサーチは、シミュレーションによって多くの到達しうる局面について探索を行うことができるため、状態数が膨大な麻雀に有効であると考えられる。

また、局所解に収束することもないため、多様な和了形を探索によって見つけることができ、状況によって戦略を変化させる必要がある麻雀に有効であると考えられる。

さらに、麻雀で和了形を目指す時、自分の捨てた牌は不要であると判断しているため、想定している和了形にその牌は必要ないはずである。したがって、和了形を探索するときに、一つ前の手牌に戻るような探索を行う必要はない。つまり、タブーサーチを用いることで、このような探索を実現することができ、効率よく和了形を見つめることができると考えられる。

4.2 評価関数

本節では、評価関数 **OPTIMUM** を構成する評価基準について述べる。今回用いた評価基準は、向聴数、受け入れ枚数、類似度と得点の 4 つである。これらの評価基準に評価の優先度を与えることで評価関数を構成する。本節では、新たに定義した類似度について述べる。

類似度は、初期手牌と探索によって見つかった和了形の手牌がどの程度類似しているかを表す値である。この値は初期手牌と和了形の手牌を比較し、変化した牌をツモによって獲得する確率をすべて乗算することによって得られる。プレイヤー i が、ある牌 x をツモによって獲得する確率 $P(i, x)$ は、プレイヤー i の手牌に存在する牌 x の枚数 $N(i, x)$ 、場に見えている牌 x の枚数 $T(x)$ を用いて、式 (8) で表される。

$$P(i, x) = \frac{4 - N(i, x) - T(x)}{136 - \sum_i \sum_x N(i, x) - \sum_x T(x)} \quad (8)$$

類似度は、変化して獲得した牌について、式 (8) を乗算することで導出できる。初期手牌と和了形の手牌に変化がなければ、類似度は最大値である 1 となり、最も類似しているとする。

4.3 麻雀へのタブーサーチの適用

麻雀へのタブーサーチの適用方法について述べる。タブーサーチのアルゴリズムを適用するために決める必要があるのは、反復回数、探索中に生成する近傍値とタブーリストの保持数である。また、打牌の決定方法についても述べる。

反復回数は、初期手牌から和了形を想定する時に、特殊な手牌である染め手などを目指す場合を除けば、それほど多くの入れ替えは必要がないと考えられる。また、反復回数を少なくすれば、アルゴリズムの実行回数を増加させることができるため、多様な和了形を見つめることができ、より良い解を見つめられると考えられる。この値は両局面にて同値とし、実験で決定する。

次に、近傍値について述べる。近傍値は、変化前の状態と類似している必要があり、例えば、巡回セールスマン問題における近傍値は現在の町の隣の町を指す。本論文の麻雀における近傍値は、現在の手牌から、任意の 1 枚の牌を残っている牌のどれかに変化させることで得ることができる。探索によって見つかった近傍値を評価し、最も評価値が高い値に遷移する。タブーリストには、現在の手牌と最適な近傍値を記録しておき、この遷移は行わないようにする。麻雀では、和了形に近づけば近づくほど、有効牌が減少し、受け入れ枚数が少なくなる。そのため、近傍値の生成数を少なくすると、和了形を探索によって見つけられないことが多くなり、得点の評価値の意味が薄くなる。したがって、麻雀において、生成数はある程度大きくする必要があるのである。

タブーリストの保持数はアルゴリズム実行ごとに全ての遷移を記録するため 6 とする。また、関連論文 [8] より、タブーリストの保持数は、実験的に 5 から 12 の値が良いとされていることも理由の一つである。

最後に、最適解を見つけた後、打牌を決定する必要がある。打牌を決定するために、初期手牌の各牌に対して、評価値として初期選択回数と平均得点を導入する。初期選択回数は、1 回目の遷移で各牌が何回打牌されたかを表す値である。平均得点は、探索によって和了することができた場合、1 回目の反復で打牌された牌に対して得点を加算しておき、初期選択回数で割ることで得られる。以上の評価値を用いて、初期手牌に存在するが、最適解には存在しない牌の中から最適な打牌を決定する。

アルゴリズムを以下に示す。以下のアルゴリズムを、設定した実行時間内で可能な限り行い、行動を決定する。提案手法において、式 (1) の x はプレイヤーの手牌を指してお

り、タブーサーチを使用することで、全ての手牌の組み合わせから、最も良い組み合わせを4つの評価基準を用いて探索する。

- (1) 初期手牌を最適解とする。反復回数を $k = 0$ とし、タブーリスト T を空集合とする。
- (2) 反復回数を $k = k + 1$ とし、現在の手牌から牌を1枚選択し、山に残っている可能性のある牌と入れ替える。そして、変化後の手牌を評価する。
- (3) 変化後の手牌が最適解よりもよい評価値を持てば、最適解を更新する。その後、現在の手牌を変化後の手牌とする。
- (4) 反復回数が規定回数に達する、または、探索時間が設定値に達した場合、停止する。もしそうでなければ、タブーリスト T を更新し、ステップ2に戻る。タブーリストには、現在の手牌と一つ前の手牌が記録される。

4.4 打牌局面の設定

打牌局面は、自分が牌をツモで獲得し、手牌が14枚ある状態のことを指す。この手牌を初期手牌として探索を開始する。

アルゴリズムの実行回数は特に定めず、ルールでの打牌時間が4秒のため、4秒間の探索を行い最適な打牌を決定する。近傍値の生成数は、実験にて決定する。

打牌選択の評価の優先度は、向聴数、受け入れ枚数、類似度、得点の順である。このような評価順とした理由は、和了することを優先したためである。

打牌を決定する際の優先度は、初期選択回数、平均得点の順である。これは、打牌選択の評価と同様で、和了することを優先したためである。最適解が二つ以上存在する場合は、ランダムでどれかを打牌する。

4.5 鳴き局面の設定

鳴き局面は、相手が打牌を行い、鳴きを行うことによって向聴数を小さくできる状態を指す。この手牌を初期手牌として探索を開始する。

アルゴリズムの実行回数は特に定めず、最大2秒間の探索を行い、鳴きを行うかどうかと最適な打牌を決定する。最大2秒間としているのは、鳴きを行う際に構成可能な面子の候補が最大で順子3種類と刻子1種類の計4種類あり、各探索時間を0.5秒としているためである。槓子は、構成しても向聴数を小さくできないため、本研究では用いない。

近傍値の生成数は打牌選択局面の半分とした。このようにした理由は、探索時間が打牌選択よりも短いことと、鳴きを行ったことにより構成された面子は打牌することができないため、初期手牌の変化させる牌の候補が減り、探索が必要な状態数が減るためである。

鳴き選択の評価の優先度は、向聴数、受け入れ枚数、類似度、得点の順である。このような評価順とした理由は、

打牌選択時と同様である。見つかった最適解が和了形に到達しており、得点が0でなければ、鳴きを行う。この条件で鳴きを行う理由は、鳴きを行うことで成立しなくなる役が存在し、和了できなくなることがあるためである。

打牌を決定する際の優先度は、初期選択回数、平均得点の順である。このような評価順とした理由は、打牌選択時と同様である。鳴きを行った場合、この条件で打牌を行う。

5. 実験

5.1 実験方法

5.1.1 実験の設定

本節では、今回の実験での麻雀のルールを説明する。基本的なルールは、インターネット麻雀対戦サイト天鳳 [9] に準拠する。以下に、変更した部分を示す。

- 喰いタンあり。
- 赤ドラなし。
- 連荘なし。
- 東風戦のみで、南入しない。
- 飛びなし。
- ダブロンなし。

本研究での提案手法は、現在の得点差を考慮して探索は行わない。また、ベンチマークプレイヤーと対局実験を行うが、このプレイヤーも得点差を考慮しないため、連荘や東風戦よりも長い試合は行う必要はないと考えたため、上記のような設定とした。

実験は、CPU Dual-Core 3.20GHz、メモリ 2GBのマシン環境で行い、実装には、C++言語を用いた。

5.1.2 牌譜との打牌の一致率

5.1.2.1 実験 1

提案手法を用いて、牌譜との打牌の一致率を調べる。実験に使用した牌譜は、インターネット麻雀対戦サイト天鳳 [9] の鳳凰卓の牌譜である。鳳凰卓で対戦を行うことができるプレイヤーは全プレイヤーの上位0.1%程度であり、この牌譜との一致率を調べれば、どの程度有効な打牌の選択が行えているかがわかると考えられる。

抽出した牌譜は、本研究では赤ドラを考慮していないため、赤ドラを使用していない牌譜となっている。また、人間のプレイヤーは得点状況によって最適行動を変化させ、さらに、相手の手牌の進み具合に応じて、攻めるか、守りに入るかを決定する。しかし、本アルゴリズムではこれらの要素を考慮して捨て牌を選択していないため、最初の局で6枚の牌を捨てるまでの局面を実験対象とする。これらの条件を満たす局面を60局面、ランダムで抽出する。

実験内容は、選択した局面を入力として与え、実際に行われた打牌の順位を出力することで行う。タブーサーチでの探索には、ランダムな要素が存在するため、毎回同じ順位を出力するとは限らない。したがって、この操作を各局面に対して50回行う。1位に選ばれた回数が過半数を超え

れば、1位での一致、1位または2位に選ばれた回数が過半数を超えれば、2位での一致、以下同様に評価する。また、反復回数と生成数のパラメータの決定を行うため、反復回数を6に固定し生成数を変化させた場合と、生成数を10に固定し、反復回数を変化させた場合で実験を行った。

5.1.2.2 実験2

提案手法を実際に対局で使用するために、実験1と異なり、各局面に対する最適な打牌を一度の探索で決定する。

抽出した牌譜は、実験1の条件に加え、立直が行われるまでの局面を選択している。これらの条件を満たす局面を125局面使用する。

実験内容は、選択した局面を入力として与え、タブーサーチでの探索を実行時間中に可能な限り行うことで打牌候補の順位を決定し、1位の打牌を出力することで実験を行った。

5.1.3 ベンチマークプレイヤーとの対局実験

提案手法を用いて作成したプレイヤーとベンチマークプレイヤー同士で対局を行い、その性能を評価する。実験の設定は、5.1.1で記載した通りである。対局は、作成したプレイヤー2人とベンチマークプレイヤー2人の計4人で行い、同じプレイヤーは向かい合うように設置する。また、同じ牌の集合に対して、親を2種類のプレイヤーに振り分けて、2回行う。評価は平均順位と和了率で行う。

5.1.3.1 作成したプレイヤー

作成したプレイヤーは、提案手法を用いて打牌のみを行うプレイヤー（以下、*M_I*）と提案手法を用いて打牌と鳴きを行うプレイヤー（以下、*M_II*）である。

共通のルールとして、両プレイヤーは、立直を宣言できる打牌をした場合、必ず立直を行う。

5.1.3.2 ベンチマークプレイヤー

ベンチマークプレイヤーは2つ作成し、1つは、向聴数を小さくする可能性のある牌の集合からランダムで打牌を行うプレイヤー（以下、*B_I*）である。もう一つは、役牌を鳴いた後、向聴数を小さくする鳴きを行うプレイヤー（以下、*B_II*）である。*B_II*は提案手法を用いて打牌を行う。

共通のルールとして、両プレイヤーは、立直を宣言できる打牌をした場合、必ず立直を行う。

5.2 実験結果

5.2.1 牌譜との打牌の一致率

5.2.1.1 実験1

反復回数を6に固定し、生成数を5、10、15と変化させた場合の一致率を表1に、生成数を10に固定し、反復回数を4、6、8と変化させた場合の一致率を表2に示す。捨て牌候補は、最大で6位まで出力された。

表1と表2の結果から、反復回数と生成数のパラメータを変化させても大きな一致率の変化は見られなかった。今後、改善を加えることで、アルゴリズムの計算量が増加す

表1 反復回数を変更した場合の一致率 (%)

| 反復回数/順位 | 1 | 2 | 3 | 4 | 5 | 6 |
|---------|------|------|------|------|------|------|
| 4 | 41.7 | 65.0 | 76.7 | 76.7 | - | - |
| 6 | 43.3 | 60.0 | 73.3 | 76.7 | 81.7 | 83.3 |
| 8 | 45.0 | 65.0 | 76.7 | 80.0 | 83.3 | 85.0 |

表2 生成数を変更させた場合の一致率 (%)

| 生成数/順位 | 1 | 2 | 3 | 4 | 5 | 6 |
|--------|------|------|------|------|------|------|
| 5 | 41.7 | 63.3 | 76.7 | 83.3 | 85.0 | 85.0 |
| 10 | 43.3 | 60.0 | 73.3 | 76.7 | 81.7 | 83.3 |
| 15 | 40.0 | 65.0 | 78.3 | 78.3 | 81.7 | 85.0 |

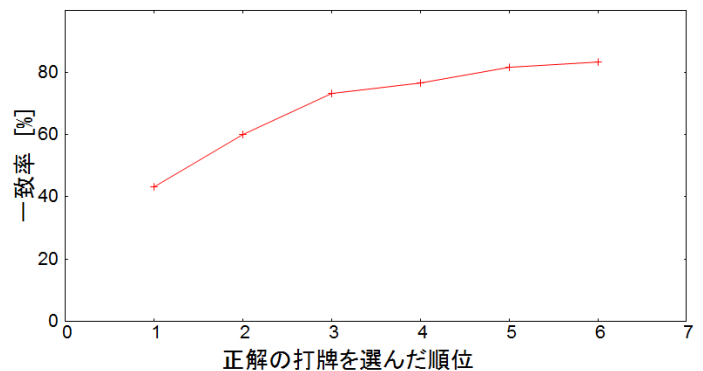


図1 打牌との一致率

ることを考慮し、生成数は1位率の最も高い10を採用したが、反復回数はできるだけ小さい値である6を対局実験で使用することとした。反復回数を4としなかったのは、5位と6位まで解候補が現れており、改善時にこの要素が使用できる可能性を考慮したためである。

反復回数を6、生成数を10とした場合の実験結果を図1に示す。図1から、1位での一致率は43%となり、6位までの一致率は83%となった。

5.2.1.2 実験2

実験1の結果から、反復回数を6、生成数を10として実験を行った。結果を表3に示す。解の候補に最適な打牌がほぼ挙がっているが、序盤の局面では、1位に最適な打牌を選べていないことが多かった。しかし、後半の局面では、最適な打牌を1位に選べる傾向にあった。ただし、染め手のような特定の役をプレイヤーが目指している局面があり、タブーサーチを用いた探索ではその点が考慮できず、探索に失敗している。そのため、10巡目から12巡目に1位率が低くなっている。また、データ数が少ないため、データ数を増やして実験を行うことが必要である。

5.2.2 ベンチマークプレイヤーとの対局実験

5.2.2.1 *M_I*と*B_I*の対局

最初に、*M_I*と*B_I*で116戦、対局を行った。結果を表4に示す。平均順位には、ウェルチのt検定により、有意水準5%で有意な差があった。表4から、*M_I*は*B_I*に比べて、平均順位も和了率も上回り、有意に強いプレイヤー

表 3 打牌の一致率

| 巡目 | 1位率 [%] | データ数 | 候補への出現率 [%] |
|----|---------|------|-------------|
| 1 | 30 | 10 | 90 |
| 2 | 30 | 10 | 90 |
| 3 | 40 | 10 | 80 |
| 4 | 20 | 10 | 80 |
| 5 | 40 | 10 | 80 |
| 6 | 50 | 10 | 70 |
| 7 | 70 | 10 | 90 |
| 8 | 80 | 10 | 80 |
| 9 | 90 | 10 | 90 |
| 10 | 50 | 8 | 50 |
| 11 | 50 | 6 | 50 |
| 12 | 50 | 6 | 100 |
| 13 | 75 | 4 | 75 |
| 14 | 100 | 3 | 100 |
| 15 | 66.67 | 3 | 66.67 |
| 16 | 100 | 3 | 100 |
| 17 | 100 | 2 | 100 |

表 4 M_I と B_I の対局実験 (%)

| プレイヤー | 平均順位 | 和了率 |
|-------|------|-------|
| M_I | 2.16 | 52.80 |
| B_I | 2.81 | 31.47 |

表 5 M_I と M_{II} の対局実験 (%)

| プレイヤー | 平均順位 | 和了率 |
|----------|------|-------|
| M_I | 2.66 | 33.09 |
| M_{II} | 2.28 | 61.03 |

表 6 M_{II} と B_{II} の対局実験 (%)

| プレイヤー | 平均順位 | 和了率 |
|----------|------|-------|
| M_{II} | 2.34 | 50.00 |
| B_{II} | 2.63 | 42.39 |

であることがわかった。

5.2.2.2 M_I と M_{II} の対局

次に、 M_I と M_{II} で 34 戦、対局を行った。結果を表 5 に示す。平均順位には、ウェルチの t 検定により、有意水準 5% で有意な差があった。表 5 から、 M_{II} は M_I に比べて、平均順位も和了率も上回り、有意に強いプレイヤーであることがわかった。

5.2.2.3 M_{II} と B_{II} の対局

最後に、 M_{II} と B_{II} で 46 戦、対局を行った。結果を表 6 に示す。平均順位には、ウェルチの t 検定により、有意水準 5% で有意な差があった。表 6 から、 M_{II} は B_{II} に比べて、平均順位も和了率も上回り、有意に強いプレイヤーであることがわかった。

6. 考察

6.1 牌譜との打牌の一致率について

表 1 と表 2 から、パラメータを変化させても一致率の大

きな変化は見られなかった。これは、今回のテストデータが序盤の局面に限ったものであり、似た局面も存在したためと考えられる。また、探索時間が 4 秒になっており、この時間が十分に長く設定されており、類似した最適解に最終的には到達している可能性も考えられる。

図 1 から、牌譜との打牌の一致率は最大で 83% となっていた。したがって、探索によって見つかった和了形には、正解の打牌によって捨てられた牌が含まれていないことが多く、序盤の局面に対して、有効な和了形であることが多かった。しかし、1 位での一致率は 43% となっており、候補に挙げることができていても、正解の打牌を選択できていないことが多かった。この理由について、考察を行う。

まず、孤立牌の優先度が低いことが挙げられる。選択した局面の中には、面子候補の形が悪いときに孤立牌を残し、面子候補を崩す場面が見られた。しかし、提案アルゴリズムは手牌の中から 1 枚の牌をランダムに変化させ、生成された手牌の中から向聴数の小さい手牌に遷移を行う。したがって、面子候補から牌を 1 枚変化させ、別の面子候補を生成しても、和了には近づいていないため、遷移する可能性は低い。一方、孤立牌を変化させ、面子候補を生成した場合、面子と面子候補の合計が 4 個以下であれば、向聴数を小さくすることができるため、遷移する可能性は高い。以上の理由から、面子候補を崩す可能性が低い点が、このような結果となった理由の一つであると考えられる。

次に、孤立牌が複数存在する局面の存在が挙げられる。提案アルゴリズムでは、孤立牌は上記の理由で最適解になりやすいが、複数存在する場合は選択頻度が分散してしまい、1 位で過半数に到達しないことがあった。その結果、1 位での一致率はあまり高くないが、2 位、3 位になるにつれて、一致率が大きく向上した。

最後に、牌譜のプレイヤーの決め打ちの存在が挙げられる。手牌が与えられた時に、プレイヤーが最終形に含まれる役を想定し、和了から遠ざかったとしても必要のない牌を打牌している局面が存在した。例えば、染め手を狙う場合、ある 1 種の数牌を除く別の数牌は、和了形には必要ないため、優先的に打牌されてしまう傾向がある。提案アルゴリズムでは、得点よりも類似度を優先しているため、このような場合には、プレイヤーが想定している和了形を探索では見つけられず、打牌が一致しなかったと考えられる。

改善案として、面子候補や孤立牌に、重要度に応じて重みを与えるという手法が挙げられる。例えば、数牌であれば、1 の牌よりも 5 の牌の方が順子を構成できる可能性が高い。そのため、近傍値を生成する時に、1 の牌を変化しやすくすることで選択頻度の偏りを避けることができると考えられる。

6.2 対局実験について

6.2.1 M_I と B_I の対局

表4から、 M_I は B_I よりも有意に強いプレイヤーとなった。 B_I は、向聴数を小さくする可能性のある牌の集合からランダムで打牌を行うプレイヤーである。このプレイヤーよりも強いプレイヤーとなった理由として、 M_I は向聴数を小さくする可能性のある牌の中で、平均的に受け入れ枚数が多くなる打牌を行っていると考えられ、その結果、和了率が B_I よりもよくなったと考えられる。麻雀は、和了することが攻めと同時に守りにもなっているため、和了率が高くなったことで、平均順位も大きく上回ったと考えられる。

6.2.2 M_I と M_{II} の対局

表5から、 M_{II} は M_I よりも有意に強いプレイヤーとなった。 M_{II} は、 M_I に鳴きを追加したプレイヤーである。鳴きは、行うことによって、役が一つも成立しないことがあり、向聴数が小さくなる場合であっても、鳴かないほうが良いことも多い。しかし、 M_{II} は鳴きを追加したことで、和了率も大きく向上し、有意に強くなったことから、提案手法における鳴きの行動決定は有効であることがわかる。また、この結果から、鳴きを追加することは、プレイヤーの実力の向上につながることもわかった。

6.2.3 M_{II} と B_{II} の対局

表6から、 M_{II} は B_{II} よりも有意に強いプレイヤーとなった。 B_{II} は役牌を鳴いてから、向聴数を小さくする場合に鳴きを行うプレイヤーである。この実験結果から、 M_{II} は役の確定しない鳴きも行っており、和了形に到達していると考えられる。提案アルゴリズムでは、探索によって見つかった最も類似した最適解から点数を得られるならば、鳴きを行う。したがって、役の確定しない鳴きを行った場合でも、無理なく和了形に到達できていると考えられる。また、鳴いた後の打牌選択も通常と同様に、得点よりも類似度の優先度を高くしていたが、和了することができていたため、鳴きの決定を行った時と同様の和了形を探索によって見つけることができていると考えられる。

7. まとめ

将棋やオセロといった2人零和確定完全情報ゲームにおいて、トップ層の人間と同等の実力を持つプレイヤーは実現されているが、ポーカーや麻雀といった不確定不完全情報ゲームで実現することは難しい。

そこで、本論文では、トップ層の人間に勝る麻雀プレイヤーを実現するために、多人数不完全情報ゲームにおける最適行動の探索について検討した。ここでは、知識を用いずに最適な行動を決定することを目的として、シミュレーションによって最適解の探索を行うタブーサーチを用いる手法を提案した。本研究では、打牌選択の局面と鳴きの局面でタブーサーチを適用した。

提案手法を実現し、牌譜との打牌の一致率実験とベンチ

マークプレイヤーとの対局実験を行った。一致率実験によって、序盤の局面に対して、有効な和了形を最大85%で見つけられているという結果が得られた。また、対局実験によって、提案手法の打牌と鳴きがベンチマークプレイヤーよりも優れているという結果が得られた。これらの結果から、タブーサーチによる最適解の探索が麻雀において有効である可能性を示した。したがって、提案手法が、例えば、ポーカーのようにプレイヤーが手札のカードを入れ替える、麻雀と行動が類似しているゲームにおいて有効である可能性が考えられる。

本研究の一致率実験では、最大85%の一致率が得られていたが、1位での一致率は最大で45%となった。したがって、考察でも述べたように、孤立牌に重みを付ける等の評価基準の追加によって、1位での一致率を向上させることが今後の課題である。

また、今回作成したプレイヤーは、降りを行わない。麻雀では、どのような手順で打牌を行っても、和了に到達できない場合がある。したがって、和了できない局面を判断し、相手にロンをさせられないような打牌を行うように、降りをプレイヤーに追加する方法を提案することも今後の課題である。

さらに、勾配法などの他の最適化手法と比較することも必要である。

最後に、上記課題を克服したプレイヤーに対して、人間のプレイヤーと対局実験を行い、性能を評価することも今後の課題である。

参考文献

- [1] N. A. Risk and D. Szafron, "Using counterfactual regret minimization to create competitive multiplayer poker agents", Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems, volume 1-Volume 1, International Foundation for Autonomous Agents and Multiagent Systems, pp. 159-166, 2010.
- [2] 水上直樹, 中張遼太郎, 浦晃, 三輪誠, 鶴岡慶雅, 近山隆, "降りるべき局面の認識による1人麻雀プレイヤーの4人麻雀への適用", 第18回ゲームプログラミングワークショップ, pp. 1-7, 2013.
- [3] 水上直紀, 鶴岡慶雅, "牌譜を用いた対戦相手のモデル化とモンテカルロ法によるコンピュータ麻雀プレイヤーの構築", 第19回ゲームプログラミングワークショップ, pp. 48-55, 2014.
- [4] 水上直紀, 鶴岡慶雅, "期待最終順位に基づくコンピュータ麻雀プレイヤーの構築", 第20回ゲームプログラミングワークショップ, pp. 179-186, 2015.
- [5] 三木理斗, 三輪誠, 近山隆, "木カーネルを用いたSVMによる麻雀打ち手の順位学習", 第13回ゲームプログラミングワークショップ, pp. 60-66, 2008.
- [6] 三木理斗, 近山隆, "多人数不完全情報ゲームにおける最適行動決定に関する研究", 修士論文, 東京大学, 2010.
- [7] 高橋翔悟, 六井淳, "麻雀における行動戦略に関する研究", 第11回情報科学技術フォーラム, pp. 280-284, 2012.
- [8] F. Glover, "Tabu Search, Part I", ORSA Journal on Computing, Vol. 1, No. 3, pp. 190-206, 1989.
- [9] 角田真吾, 天鳳, <http://tenhou.net/>, 2016.