麻雀 1 局の目的に応じた抽象化と価値推定からなるプレイヤの開発

栗田 萌[1]　保木 邦仁[2]

概要：本研究では、麻雀の 1 局を複数の目的（アガリ、形式テンパイ、降り）に応じて 1 人麻雀として抽象化し、機械学習法により期待順位点を推定して麻雀プレイヤを開発する方法を提案する。ソースコードが公開されている麻雀プレイヤ manue と 1500 半荘以上の対局実験を行い、平均順位 2.193 ± 0.026 の結果を得た。

Development of Mahjong Player on the basis of *Kyoku* Abstraction for Multiple Goals and Value Inference

Moyuru Kurita[1]　Kunihito Hoki[2]

Abstract: In this research, we propose methods to develop a mahjong player by using single-player-mahjong models as abstraction of real *kyoku* with a specific goal (*agari*, *keishiki tenpai*, or *ori*), and inference of expected final points on the basis of machine learning techniques. We have carried out over 1500 game play vis-à-vis manue whose source codes are available online, and obtained an estimate of expexted final ranking of the player at 2.193 ± 0.026.

1. はじめに

ゲームは勝敗が明確に定まることから古くから人工知能 (AI) の研究の対象とされ、多くのゲームにおいて人間のトッププレイヤと互角以上に戦う AI が開発されてきた。チェス、将棋、囲碁のような偶然の要素がゲームの進行に本質的な影響を与えない二人ゼロ和確定ゲームについては、ゲーム木の探索手法や局面の評価値を算出する手法などの発展を背景に、AI が世界チャンピオンを破るに至った [*1][1], [2]。不確定ゲームも二人ゼロ和であれば、バックギャモンや 2 人ポーカーなどで人間のトッププレイヤ以上の強さを持つ AI が開発されている [3], [4]。一方で、三人以上のゲームに関する AI の研究は、二人ゼロ和ゲームのそれほど発展しているとは言い難い。

麻雀も、通常プレイヤ数が 4 であり、現状トッププレイヤ以上の強さを持つ AI が開発されたとは言い難い状況で、強化学習のように AI に大量の自己対局を行わせてその結果を学習する方法も確立されていない。実際、麻雀には強いコンピュータのプレイヤの構築を困難にする複数の性質があると考えられる。1 つめは非常に大きなゲーム木を持つという性質である。麻雀ではプレイヤが行動を選択する前に、常にツモや他家（自分以外のプレイヤ）の打牌などに偶然の要素が影響を与え、牌種も 30 種類以上存在するためゲーム木のサイズが膨大になってしまう。2 つめは、複数の局からなる半荘で勝つという長期的な目的を達するためには、各局で短期的な目的を適切に設定しなければならないという性質である。麻雀では一局の中でアガリを行うことを目指すべきか、放銃しないことを目指すべきかの選択を適切に行うことが要求され、これらの各局の状況を統一的なゲーム木として抽象化することは困難である。3 つめは局を終了させるアガリに関するルールが複雑であるという性質である。麻雀には数十種類の役が存在し、それぞれで異なるアガリ点数が定められている。そのため、探索を行わない静的評価関数を麻雀で実現することを試みる場

[1] HEROZ 株式会社 mkmjai1@gmail.com
[2] 電気通信大学 k.hoki@uec.ac.jp
[*1] http://www.ipsj.or.jp/50anv/shogi/20151011.html (2017)

合、各役に対応した特徴量を設計する必要があり、これはAI 開発者に麻雀の技術を要求するだけでなく、前述の困難も合わさって特徴量の次元が大きくなる。4 つめはプレイヤの実力評価が難しいという性質である。麻雀では正しい選択をしたかによらず運が良いプレイヤが勝ってしまうことが珍しくない。結果的に 2 つの異なる AI の強さを比較する段階で、多くの数の対局を行う必要性が生じる。実際、あるプレイヤが N 回の対局を行った場合の平均順位の標準誤差はおよそ $\sqrt{1.25/N}$ 程度であり、$N = 1000$ として 0.035 である [5]。天鳳の鳳凰卓では、最上位プレイヤの平均順位が 2.40 程度であり、このようなレベルに達した 2 プレイヤ間で実力を比較するためには多くの試行が必要である *2。このような困難は、対局結果にもとづいてパラメタ一式の更新や調整を行う強化学習において特に顕著にあらわれるであろう。前述のとおり、麻雀の評価関数の次元は大きくなりやすく、各パラメータを適切に収束させるために必要なデータの量は、パラメータの数だけでなく、麻雀そのものが持つ運の要素の寄与で大きくなる。これらの困難があるためか、評価関数のパラメータを対局の結果から学習する手法は、強い AI を作る目的では現在成功していない。

我々は先行研究において、麻雀の手役作りに関して 1 つめと 2 つめの問題を解決するため、1 人麻雀のゲーム木を、サイズの小さい有向非巡回グラフへ抽象化する手法を提案した [6]。本研究では、放銃確率の推定をもとに [7]、降りも考慮したゲーム木として麻雀の 1 局を抽象化することで 1 つめと 2 つめの問題の解決を図る。また、3 つめと 4 つめの性質により生じる、パラメータ一式の調整に要するデータ数が膨大になる困難についても解決を図る。具体的には、抽象化したゲーム木の探索と局面評価の学習を組み合わせて、少ない特徴量から対局結果を学習する手法を提案する。本研究では学習に天鳳の鳳凰卓の牌譜を用いているが、最終的には AI が自己対局を繰り返しその結果に基づいて学習を行うことができる枠組みを確立することを目的としている。

2. 先行研究

麻雀 AI の先行研究は、アガリを目指す 1 人麻雀を用いるものが多い。代表的な研究として、パーセプトロンを用いて上級者の選択との一致を目指すものや [8]、モンテカルロ法やタブーサーチを用いてゲームの試行や木探索をするものがある [9], [10]。一方で、アガリを目指すだけでは中級者以上の実力を実現することは不可能と考えられるが、降りの戦略を含めた麻雀 AI の研究は例が少ない。そのような研究の例として、5000 以上の局面に対して降りるべき局面を開発者が手動でラベル付けし、その結果を元に機械学習を行う水上らの研究がある [11]。また、この手法とモンテカルロ法による試行を併用することで、中級者以上の実力を持った AI が実現されている [12]。降りるべき局面の判別に関しては、ラベル付けを人手で行うにはコストがかかるという動機から、半教師あり学習を用いて判別する研究も報告されているが [13]、包括的な麻雀 AI に利用されているという報告はされていない。

1 人麻雀はその有効性が示されているものの、いくつかの課題が存在する。まず、点数状況に応じたアガリを目指すことができていないことが挙げられる。手牌からアガリ点数を予測するモデルを用いた強化学習を行った水上らの研究で、既存のプログラム（点数状況に応じたアガリを目指すことができないもの）に勝ち越すことができなかったと報告されている [14]。また、機械学習に用いる特徴は麻雀の専門知識に基づき設計されるものが多く、AI 開発の敷居が高いことも挙げられる。

3. 提案手法概要

本研究では、麻雀 1 局の目的に応じてゲーム木を複数に抽象化し、その探索結果を用いて局面や選択の価値を推定する手法を提案する。全ての抽象化において、他家の手番は全て偶然プレイヤの手番として置き換えられ、プレイヤが偶然プレイヤと自分（意思決定を行うプレイヤ）のみとなるため、これらを 1 人麻雀と呼ぶ。自分の手牌や他家の状態で大まかに場合分けを行い、1 人麻雀の結果を用いる（図 1 参照）。自分の手牌のシャンテン数が大きく、他家のテンパイ確率も高くない状況では、ルールベースの打牌を行い、フーロは行わない。このルールとは、孤立字牌を対子になりにくく（既に場に見えている）、なっても価値が低いもの（刻子になっても役にならないオタ風牌）を優先的に捨てるといったものである。自分の手牌のシャンテン数が大きく、他家がリーチまたは 2 回以上のフーロを行っている場合、降りを行う。これは降り方策 1 人麻雀の結果を用いるが、その詳細は 4.2 節で説明する。自分の手牌が 1 シャンテン以下の場合は、包括方策 1 人麻雀の結果を用いる。これは、アガリ、形式テンパイ、降りなどの要素を取り入れた 1 人麻雀であり、この詳細は 4.3 で説明する。前述した条件以外の手牌では、リーチやフーロした他家がいない場合、放銃を考慮せずアガリを目指すアガリ方策 1 人麻雀の結果を用いる。この 1 人麻雀におけるゲーム木の簡略化方法は筆者らの先行研究で説明したものである [6]。リーチやフーロをした他家がいる場合は、各種の 1 人麻雀の結果を特徴とする価値推定を機械学習によって行い、利得期待値を最大化する選択を行う。

ここで本研究で用いる 4 つの 1 人麻雀について概説する。これらは麻雀 1 局においてプレイヤが取り得る各々の目的に応じたものであり、ゲーム木の構造などは異なるが、終端節点での利得の設定などで共通する部分もある。

*2 角田真吾. 天鳳. http://tenhou.net/ 2017.

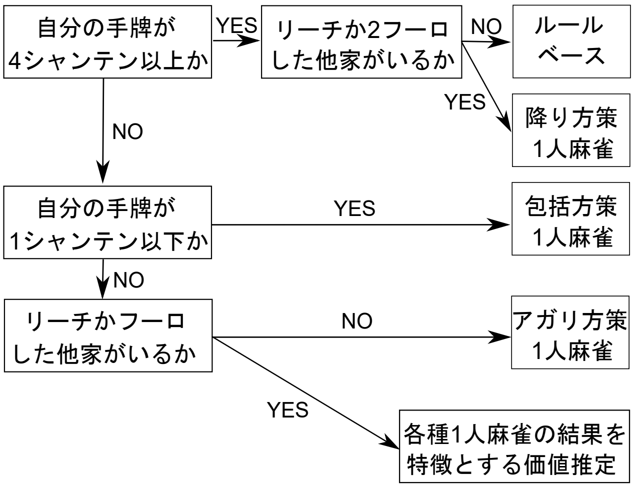

図 1　提案手法における選択のフローチャート

(1) **アガリ方策 1 人麻雀**：プレイヤは定められた打牌回数の範囲でツモとフーロを行い、アガリを目指す。プレイヤ手番の節点はツモ節点とフーロ節点が存在する。ツモ節点では、牌をツモした状態で何を打牌するか、またはアガリ宣言を行うかの選択を行い、フーロ節点では捨てられた牌をフーロするかの選択を行う。各順目においてプレイヤがツモする牌種や、フーロの対象となる捨て牌の牌種は偶然プレイヤに決定される。アガリを行った終端節点の利得は、アガリの種類（ツモアガリ、ロンアガリ）、翻、符、局などにより定まるが、麻雀のルールに基づく点数そのものではなく、局終了時の順位点推定値を用いる。アガリをしない状態で定められた打牌回数に達した終端節点での利得は手牌によらないものとする。放銃や他家のアガリは考えない。

(2) **形式テンパイ方策 1 人麻雀**：プレイヤは定めれらた打牌回数の範囲でツモとフーロを行い、形式テンパイを目指す。アガリ方策 1 人麻雀との違いは、プレイヤはアガリ宣言を行うことができないことと、定められた打牌回数に達した終端節点での利得が手牌に依存することである。プレイヤの目的は形式テンパイを狙うことのみであるため、終端節点で手牌が形式テンパイの場合の利得は 1 で、そうでない場合の利得は 0 とする。

(3) **降り方策 1 人麻雀**：プレイヤは定められた打牌回数に達するまで打牌を行うが、ツモやフーロは行わない。したがって、打牌回数は手牌枚数より小さい値である。プレイヤの打牌した牌種が放銃になるかどうかは偶然プレイヤに決定される。この時の放銃確率は打牌回数に依存しないが、一度放銃とならなかった牌の放銃確率は 0 とする。また、プレイヤが 1 回打牌し放銃しない場合、偶然プレイヤにより一定の確率でゲームが終了する。これは他家のツモアガリ、もしくは他家から他家のロンアガリに対応している。終端節点の利得は、放銃した節点における牌種ごとに異なる利得、偶然プレイヤによりゲームが終了する節点の利得、定められた打牌回数に達した場合の利得が存在する。

(4) **包括方策 1 人麻雀**：プレイヤは定められた打牌回数の範囲でツモとフーロを行い、放銃を避けながらアガリまたは形式テンパイを目指す。アガリを行った終端節点の利得はアガリ方策 1 人麻雀と同じである。また、定められた打牌回数に達した終端節点の利得は手牌に依存する。この利得は、形式テンパイで流局した場合とそうでない場合それぞれでの順位点推定値を用いる。プレイヤが打牌した後には偶然プレイヤによってそれが放銃となるか判定される。この確率は牌種と巡目に依存し、放銃した節点の利得は牌種に依存する。放銃とならない場合は、次にフーロの選択を行うまでの進行は以下のようである。はじめに、プレイヤが降り方策を選択するか決定する。降り方策を選択した節点は終端節点として表され、その利得は降り方策 1 人麻雀の結果をもとに算出する。降り方策を選択しない場合、偶然プレイヤによって他家のアガリに対応する終局をするかの選択が行われる。終局とならない場合、偶然プレイヤがフーロの対象となる牌を選択する。

これらの 1 人麻雀において、考慮されている要素とされていない要素を表 1 にまとめた。4 つの 1 人麻雀のうち、アガリ方策 1 人麻雀と形式テンパイ方策 1 人麻雀は、先行研究で提案した 1 人麻雀のモデルの終端節点の利得を変更し部分的に簡略化したものと考えることができる [6]。降り方策 1 人麻雀は、ツモやフーロを行わないため、他の 1 人麻雀と比べてゲーム木のサイズが小さく扱いが容易である。包括方策 1 人麻雀は、麻雀の多くの要素を取り込んでいるため、このゲーム木における期待値最大化方策に従った選択が妥当なものになる場合が多い。一方で、例えば放銃をほとんど気にする必要が無い 1 局の序盤などは単純にアガリ方策 1 人麻雀の選択をした方が良い場合もある。実際にどのような状況でどの 1 人麻雀を利用するかを決めるフローは著者の経験則に基づいて行われていて、改善の余地がある。

表 1　各種 1 人麻雀において考慮されている要素

	ツモ・フーロ	アガリ	形テン	放銃
アガリ方策	○	○	×	×
形式テンパイ方策	○	×	○	×
降り方策	×	×	×	○
包括方策	○	○	○	○

4. 各種 1 人麻雀について

この章では、前章で概説した 1 人麻雀の詳細について説明する。はじめに 1 人麻雀の終端節点の利得を定めるため

に、局の終わりの価値（1 人麻雀プレイヤの期待順位点の何らかの推定値）について、その計算方法を説明する。次に、降り方策 1 人麻雀と包括方策 1 人麻雀について、定式化と利得を最大化する方策を説明する。

4.1 1 人麻雀の終端節点の利得

この節では、後の抽象化した 1 人麻雀で利用する価値について定義と計算方法を説明する。

はじめに、麻雀の 1 局が終わり方が確定した場合の価値を定める。麻雀の 1 局の終わり方は、アガリと流局の 2 種類に分類でき、その際の各プレイヤの点数の収支はルールにより定められている。1 局の終わり方がアガリである場合、プレイヤ i がプレイヤ j から翻 x_{han}, 符 x_{fu} の手でアガリをした場合の価値を $R^{ij}_{\text{agari}}(x_{\text{han}}, x_{\text{fu}})$ とする。ここで、ツモアガリは $i = j$ であり、ロンアガリは $i \neq j$ である。また、1 局の終わり方が流局である場合、価値は各プレイヤの聴牌フラグを用いて表すことが可能で、これを $R_{\text{ryuukyoku}}(l_1, l_2, l_3, l_4)$ と書く。ここで l_i はプレイヤ i がテンパイしている場合に 1 で、そうでない場合は 0 である。また、自分のプレイヤ番号は $i = 1$ とする。これらの価値は多クラスロジスティック回帰による順位確率の推定から求まる [6]。

続いて、1 局の終わり方に関する確率をいくつか定義し、その推定法を説明する。自分がアガリをできない場合に流局する確率を $p_{\text{ryuukyoku}}$ とする。推定は、自分の打牌回数に関して場合分けを行い、以下の特徴量ベクトルを用いたロジスティック回帰により行う。

$$
\begin{aligned}
&\phi_{\text{ryuukyoku}} \\
&= \begin{pmatrix} 1 \\ \text{2 人以上の他家がリーチしているフラグ} \\ \text{1 人の他家のみがリーチしているフラグ} \\ 1 - \prod_{i=2}^{4}(1 - p^i_{\text{tenpai}}) \end{pmatrix} \qquad (1)
\end{aligned}
$$

ここで、p^i_{tenpai} はプレイヤ i のテンパイ確率であり、著者らの先行研究の手法で推定する [7]。また、4 行目のテンパイ確率に関する積は、リーチしていない他家についてのみ行う。また、自分がアガリをせずさらに流局しない場合、他のプレイヤがアガリをすることになるが、その確率を $p^i_{\text{agari-other}}$ とする。他のプレイヤ i は下家、対面、上家のいずれかであり、この確率の推定は 3 クラスのロジスティック回帰により行う。特徴量のベクトルは 1、他家 3 人のリーチフラグ、他家 3 人のテンパイ確率の 7 次元である。さらに、他家 i がアガリをし、さらにそれが自分の放銃でない場合に、それが他家 j からのアガリである確率を $p^{ij}_{\text{agari-from}}$ とする。この値は、経験則に基づいて

$$
p^{ij}_{\text{agari-from}} = \begin{cases} \frac{5}{9} & i = j \\ \frac{2}{9} & i \neq j \end{cases} \qquad (2)
$$

とする。流局した場合に他家 i が聴牌している確率を $p^i_{\text{ryuukyoku-tenpai}}$ とする。他家 i がリーチしていない場合、この推定はロジスティック回帰に基づいて行い、特徴量のベクトルは、$\phi_{\text{ryuukyoku-tenpai}} = (1, p^i_{\text{tenpai}})$ とする。

最後に、後の 1 人麻雀で用いるためのいくつかの価値を定義する。他家のアガリ点に対する価値 $R^{ij}_{\text{agari}}(x_{\text{han}}, x_{\text{fu}})$ と、そのアガリの確率から他家のアガリに関する終端節点の利得が得られる。著者らの先行研究の手法を用いると、牌 h を打牌する際のプレイヤ i に対する放銃確率 $p^i_{\text{houjuu}}(h, x_{\text{han}}, x_{\text{fu}})$ を推定ことができる [7]。また、同じ手法を用いることで、プレイヤ i が牌 h をツモした際にアガリをする確率 $p^i_{\text{tsumo-agari}}(h, x_{\text{han}}, x_{\text{fu}})$ を推定することができる。これらを用いると、他家 i がプレイヤ j からアガリをする際の価値 R^{ij}_{agari} は $i \neq j$ のとき

$$
\begin{aligned}
R^{ij}_{\text{agari}} &= \frac{1}{\sum_{h \in H} n(h) \sum_{han,fu} p^i_{\text{houjuu}}(h, x_{\text{han}}, x_{\text{fu}})} \times \\
&\Bigl(\sum_{h \in H} n(h) \sum_{x_{\text{han}}, x_{\text{fu}}} p^i_{\text{houjuu}}(h, x_{\text{han}}, x_{\text{fu}}) R^{ij}_{\text{agari}}(x_{\text{han}}, x_{\text{fu}})\Bigr) \qquad (3)
\end{aligned}
$$

となる。ここで $n(h)$ は牌種 h のプレイヤから見えていない牌の枚数である。$i = j$ の時は、$p^i_{\text{houjuu}}(h, x_{\text{han}}, x_{\text{fu}})$ を $p^i_{\text{tsumo-agari}}(h, x_{\text{han}}, x_{\text{fu}})$ で置き換える。他家がアガリをし、自分が放銃しない終端接点の利得 $U_{\text{agari-no-relation}}$ は今までの結果から

$$
\begin{aligned}
&U_{\text{agari-no-relation}} \\
&= \sum_{i=2}^{4} p^i_{\text{agari-other}} \sum_{j=2}^{4} p^{ij}_{\text{agari-from}} R^{ij}_{\text{agari}} \qquad (4)
\end{aligned}
$$

と算出することができる。流局時に自分が聴牌していない終端接点の利得を $U_{\text{ryuukyoku-noten}}$、聴牌している終端接点の利得を $U_{\text{ryuukyoku-tenpai}}$ とすると、

$$
\begin{aligned}
&U_{\text{ryuukyoku-noten}} \\
&= \sum_{l_2, l_3, l_4 = 0}^{1} R_{\text{ryuukyoku}}(0, l_2, l_3, l_4) \prod_i p^i_{\text{rt}}(l_i) \\
&U_{\text{ryuukyoku-tenpai}} \\
&= \sum_{l_2, l_3, l_4 = 0}^{1} R_{\text{ryuukyoku}}(1, l_2, l_3, l_4) \prod_i p^i_{\text{rt}}(l_i) \qquad (5)
\end{aligned}
$$

と表される。ここで、$p^i_{\text{rt}}(1) = p^i_{\text{ryuukyoku-tenpai}}$ であり $p^i_{\text{rt}}(0) = 1 - p^i_{rt}(1)$ である。自分がアガリをしたり、流局時に聴牌を取ることを放棄し、放銃しないように打牌する戦略を「降り」と呼ぶ。本研究ではこの戦略を取った場合に放銃にならずに局が終わる場合、降りに成功したとする。その終端接点の利得を

$$
\begin{aligned}
U_{\text{ori-succeed}} &= p_{\text{ryuukyoku}} U_{\text{ryuukyoku-noten}} \\
&\quad + (1 - p_{\text{ryuukyoku}}) U_{\text{agari-no-relation}} \qquad (6)
\end{aligned}
$$

とする。

4.2 降り方策 1 人麻雀

麻雀では降りの技術は非常に重要な戦略である。放銃率を下げるため、基本的に安全度の高い牌から捨て、安全度の差が大きくない場合は持っている枚数が多い牌や、放銃しても安い牌から捨てていく戦略が良いとされる。ここでは、降り方策 1 人麻雀という降りを簡略化したゲームを定義し、その中で最適な戦略を求めることで実際の多人数で行う麻雀における降りの戦略を立てる手法を提案する。手組について 1 人麻雀を考えるのと同様に、降りについても問題の簡略化を行い、降り方策 1 人麻雀を以下で定義する。

- 自分の手牌には牌種 h が n_h 枚あり、各牌種 $h \in H$ ごとに、放銃確率 $p_{\mathrm{houjuu}}(h)$ と放銃時の利得 $U_{\mathrm{houjuu}}(h)$ が設定されている。
- プレイヤは自分の手番で牌を捨てる。この時、一度放銃にならなかった牌種は次も放銃にならないものとする。放銃にならない場合、一定の確率 α でゲームが終了して利得 $U_{\text{agari-no-relation}}$ を得る。終了しなかった場合、ツモは行わずプレイヤの手番に戻る。
- 自分の手牌が無くなった場合、利得 $U_{\text{ori-succeed}}$ を得てゲームが終了する。

このゲームにおいて期待利得を最大にするような打牌の順番を考える。基本的に全ての $h \in H$ に対して、$U_{\text{agari-no-relation}} > U_{\mathrm{houjuu}}(h)$ であることを想定しているため、一度放銃にならなかった牌種が手牌に残っている場合は、その牌を優先的に捨てるべきことは自明であり、最適な打牌の順番は牌種順により表現される。牌種 h が放銃にならない場合に、h を n_h 回連続で打牌する間にゲームが終端しない確率は $(1-\alpha)^{n_h}$ である。k 番目に捨てる牌種を h_k、その期待値は

$$\begin{aligned}E &= p_{\mathrm{h}}(h_1)U_{\mathrm{h}}(h_1)\\ &+ (1-p_{\mathrm{h}}(h_1))\left[1-(1-\alpha)^{n_{h_1}}\right]U_{\text{agari-no-relation}}\\ &+ (1-p_{\mathrm{h}}(h_1))(1-\alpha)^{n_{h_1}}p_{\mathrm{h}}(h_2)U_{\mathrm{h}}(h_2)\\ &+ \cdots\\ &+ (1-\alpha)^{\sum_k n_{h_k}}\left[\prod_{k=1}^{K}(1-p_{\mathrm{h}}(h_k))\right]U_{\text{ori-succeed}} \quad (7)\end{aligned}$$

と級数の形で表すことができる。ここで K は手牌に 1 枚以上含まれる牌種の数である。また、houjuu は h と略記した。これを $E(h_1,\cdots,h_K)$ と書く。これを最大化する降り順は解析的に求めることができる。この級数は

$$E(h_1,\cdots,h_k,h_{k+1},\cdots) \geq E(h_1,\cdots,h_{k+1},h_k,\cdots) \quad (8)$$

であることの必要十分条件が

$$\begin{aligned}f(h_k) &\leq f(h_{k+1})\\ f(h) &= \frac{p_{\mathrm{h}}(h)(U_{\text{agari-no-relation}}-U_{\mathrm{h}}(h))}{1-(1-p_{\mathrm{h}}(h))(1-\alpha)^{n_h}} \quad (9)\end{aligned}$$

であるという性質を持つ。したがって、$f(h)$ が小さい牌種から捨てることで、降り方策 1 人麻雀の期待値を最大化できる。$f(h)$ が $U_{\text{ori-succeed}}$ によらないのは、式 (7) において順序によらない定数項でのみ現れるためである。これは放銃率が低いもの、枚数が多いものを優先すべきという順序になっていて、直感ともあっている。以下では、この順番に並べた際に i 番目となる牌種を h_i^* と書く。この戦略を用いて降りを行った場合の放銃確率は

$$\begin{aligned}&p_{\text{ori-houjuu}}\\ &= p_{\mathrm{h}}(h_1^*) + (1-p_{\mathrm{h}}(h_1^*))(1-\alpha)^{n_{h_1^*}}p_{\mathrm{h}}(h_2^*) + \cdots \quad (10)\end{aligned}$$

と表すことができる。

後の計算で利用するため、放銃時の平均利得 $R_{\text{houjuu-average}}$ を算出する。その際の期待値は式 (7) より求めることができて、これを E_{ori} と書くと、$R_{\text{houjuu-average}}$ は

$$\begin{aligned}E_{\mathrm{ori}} &= p_{\text{ori-houjuu}}R_{\text{houjuu-average}}\\ &+ (1-p_{\text{ori-houjuu}})(1-\alpha)^{\sum_k n_{h_k}}U_{\text{ori-succeed}}\\ &+ (1-p_{\text{ori-houjuu}})\left(1-(1-\alpha)^{\sum_k n_{h_k}}\right)U_{\text{agari-no-relation}} \quad (11)\end{aligned}$$

より求めることができる。α の値については、本研究では暫定的に 0.1 とする。また、放銃に関する値については

$$\begin{aligned}p_{\mathrm{houjuu}}(h) &= \sum_{i,x_{\mathrm{han}},x_{\mathrm{fu}}} p^i_{\mathrm{houjuu}}(h,x_{\mathrm{han}},x_{\mathrm{fu}})\\ U_{\mathrm{houjuu}}(h) &= \frac{1}{p_{\mathrm{houjuu}}(h)}\\ &\sum_{i,x_{\mathrm{han}},x_{\mathrm{fu}}} p^i_{\mathrm{houjuu}}(h,x_{\mathrm{han}},x_{\mathrm{fu}})R^{i1}_{\mathrm{agari}}(x_{\mathrm{han}},x_{\mathrm{fu}}) \quad (12)\end{aligned}$$

と計算する。

4.3 包括方策 1 人麻雀

この節では放銃や他家のアガリ、降りを考慮に入れてアガリや形式テンパイを目指す包括方策 1 人麻雀を提案する。包括方策 1 人麻雀は、先行研究で提案した 1 人麻雀に以下の 3 つの拡張を行ったものとみなすことができる [6]。

- プレイヤが打牌を行う際、順目とその時のプレイヤの手牌 q と牌種に応じた確率 $p_{\mathrm{houjuu}}(q,h,t)$ で打牌が放銃となり、利得 $U_{\mathrm{houjuu}}(q,h,t)$ が与えられてゲームが終了する。
- 偶然プレイヤがプレイヤがフーロ、ロンできる牌を選択する前に、順目とその時のプレイヤの手牌 q に応じた確率 $p_{\text{kyoku-end}}(q,t)$ によりゲームが終了し、利得 $U_{\text{agari-no-relation}}$ が与えられる。
- プレイヤは打牌を行い、それが放銃とならなかった場合、降りを選択することができて、順目と手牌に応じた利得 $U_{\mathrm{ori}}(q,t)$ が与えられてゲームが終了する。

これにより、節点の期待値は以下のように変更される。打牌を行うフェーズでは、放銃を考慮し、

$$E''(q, i, \mathrm{player_D}, t) = \max_{(q_c, i_c, a_c) \in C''(q, i, \mathrm{player_D})} [p_\mathrm{h}(q, h, t) U_\mathrm{h}(q, h, t) + (1 - p_\mathrm{h}(q, h, t)) E''(q_c, i_c, a_c, t+1)] \quad (13)$$

となる。ただし、各値の根節点に対する依存性は表記を省略した。ここで h は houjuu を意味する。他の 1 人麻雀の場合と異なり、$C''(q, i, \mathrm{player_D})$ は、降りの節点 $(q, \mathrm{null}, \mathrm{ori})$ を含み、

$$E''(q, \mathrm{null}, \mathrm{ori}, t+1) = U_\mathrm{ori}(q, t+1) \quad (14)$$

である。これにより、将来的な降りの期待値を考慮した打牌が可能となる。また、フーロを待つ偶然節点においては、それ以前に他家がアガリをし得ると考えて

$$E''(q, \mathrm{null}, \mathrm{chance_F}, t) = p_\mathrm{ke}(q, t) U_\text{agari-no-relation} + (1 - p_\mathrm{ke}(q, t)) \times \sum_{h \in H} p''_F(q, h, t) E''(q, h, \mathrm{player_F}, t) \quad (15)$$

とする。ここで、ke は kyoku-end を表す。

包括方策 1 人麻雀で新たに設定された利得や確率の値は以下の手順で求める。定められた回数の打牌を行った終端節点の利得は

$$E''(q, \mathrm{null}, \mathrm{chance_F}, N_\text{dahai-max}) = \begin{cases} U_\text{ryuukyoku-tenpai} & q \text{ がテンパイ形} \\ U_\text{ryuukyoku-noten} & \text{それ以外} \end{cases} \quad (16)$$

とする。

次に、t 回打牌を行った際に牌種 h を打牌した場合に、それが放銃となる確率 $p_\mathrm{houjuu}(q, h, t)$ とその利得 $U_\mathrm{houjuu}(q, h, t)$ について考える。先行研究において、他家 i に対する現在の放銃確率 $p^i_\mathrm{houjuu}(h, x_\mathrm{han}, x_\mathrm{fu})$ は、テンパイ確率 p^i_tenpai とテンパイ時放銃確率 $p^i_\text{tenpai-houjuu}(h, x_\mathrm{han}, x_\mathrm{fu})$ の積

$$p^i_\mathrm{houjuu}(h, x_\mathrm{han}, x_\mathrm{fu}) = p^i_\mathrm{tenpai} p^i_\text{tenpai-houjuu}(h, x_\mathrm{han}, x_\mathrm{fu}) \quad (17)$$

によって表した [7]。プレイヤが t 回めの打牌を手牌 q で行った時の放銃確率については、テンパイ確率のみが q, t に依存すると考えて

$$p^i_\mathrm{houjuu}(q, h, x_\mathrm{han}, x_\mathrm{fu}, t) = p^i_\text{tenpai-future}(q, t) p^i_\text{tenpai-houjuu}(h, x_\mathrm{han}, x_\mathrm{fu}) \quad (18)$$

とする。ここで、プレイヤが t 回打牌を行った際のリーチをしていない他家 i のテンパイ確率は、特徴 $\phi^i_\text{tenpai-future} = (1, \mathrm{logit}(p^i_\mathrm{tenpai}))$ を用いて推定する。確率の推定は、現在の打牌回数と t と手牌 q がリーチかどうかについて場合分けをして行う。これらを用いて、

$$p_\mathrm{houjuu}(q, h, t) = \sum_{i, x_\mathrm{han}, x_\mathrm{fu}} p^i_\mathrm{houjuu}(q, h, x_\mathrm{han}, x_\mathrm{fu}, t)$$

$$U_\mathrm{houjuu}(q, h, t) = \frac{1}{p_\mathrm{houjuu}(q, h, t)} \sum_{i, x_\mathrm{han}, x_\mathrm{fu}} p^i_\mathrm{houjuu}(q, h, x_\mathrm{han}, x_\mathrm{fu}, t) R^{i1}_\mathrm{agari}(x_\mathrm{han}, x_\mathrm{fu}) \quad (19)$$

と計算する。

続いて、プレイヤが t 回の打牌を行い放銃にはならなかった場合に、次の打牌をする前に他家の誰かがアガリをして局が終了する確率 $p_\text{kyoku-end}(q, t)$ について考える。これについても、特徴を

$$\phi_\text{kyoku-end} = \begin{pmatrix} 1 \\ 1 - \prod_{i=2}^{4}(1 - p^i_\mathrm{tenpai}) \end{pmatrix} \quad (20)$$

としてロジスティック回帰を行う。確率の推定は現在の打牌回数と t と手牌 q がリーチしているかについて場合分けする。

次に降り戦略を選択した終端節点の利得 $U_\mathrm{ori}(q, t)$ について考える。これは、特徴

$$\phi_\text{ori-node-houjuu}(q, t) = \begin{pmatrix} 1 \\ \mathrm{logit}(p_\text{ori-houjuu}(q, t)) \\ \text{プレイヤのフーロ回数} \end{pmatrix} \quad (21)$$

を用いて、降り選択を行って放銃する確率 $p_\text{ori-node-houjuu}(q, t)$ を推定する。ここで、$p_\text{ori-houjuu}(q, t)$ は 4.2 節の手法で計算するが、放銃確率を計算する級数 (10) 式は、捨てた牌の数が $N_\text{dahai-max} - t$ 以上になった段階で打ち切る。また、確率 $p_\text{ori-node-houjuu}(q, t)$ の推定は t によって場合分けするほか、学習データは注目したプレイヤがリーチまたは 2 回以上のフーロを行っていない 2 シャンテン以上の状態で、他家がリーチを行っている牌譜のみを用いた。これは、この状況であれば上級プレイヤの多くが降りを選択すると予想されるためである。以上より、降りの終端節点の利得を

$$U_\mathrm{ori}(q, t) = p_\text{ori-node-houjuu}(q, t) R_\text{houjuu-average} + (1 - p_\text{ori-node-houjuu}(q, t)) U_\text{ori-succeed} \quad (22)$$

とする。

5. 1 人麻雀の結果を特徴とする価値推定

この章では、今まで説明した 1 人麻雀の結果を特徴として、麻雀において価値を推定する手法を提案する。$N_\text{dahai-max}$ を、全員がフーロをせず流局する場合のプレイヤの打牌回数と設定した包括方策 1 人麻雀は、麻雀の多くの要素を取り入れて抽象化したゲームとなっているが、全ての状況でこの手法を用いることが必ずしも最良の選択になるとは限

らない。例えば、包括方策 1 人麻雀では将来的に他家のテンパイ率が高くなることを推定して安全牌を残す選択をするが、初打からこのようなことをするのは有効でない場合が多い。

この手法は、アガリ方策 1 人麻雀、形式テンパイ方策 1 人麻雀、降り方策 1 人麻雀において算出した結果を特徴として、機械学習を用いてアガリ確率、形式テンパイ確率、放銃確率を推定して、価値推定を行う手法であり、概念図を図 2 に示す。

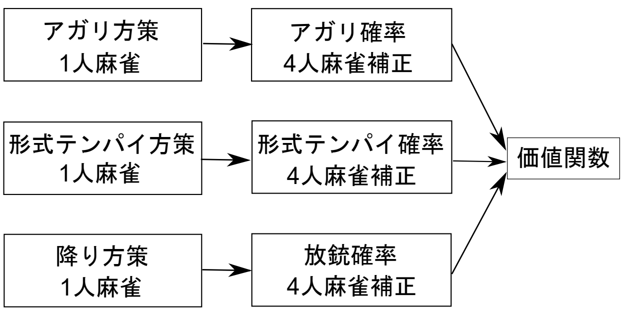

図 2 各種 1 人麻雀から価値関数を計算する仕組み

この手法を用いる場合、1 人麻雀の探索を行う段階では N_{dahai} 回の打牌を行ってアガリ出来ない終端接点の利得を、手牌がテンパイ形かどうかによらず $U_{\text{not-agari}}$ に固定する。これは、後に行う期待値計算を容易にするための処理である。

打牌回数 N_{dahai} は以下のように決める。はじめに、現在の節点から誰もアガリやフーロを行わない場合に自分が流局までに打牌する回数を $N_{\text{dahai-max}}$ とおく。探索に用いる打牌回数は $N_{\text{dahai-max}}$ より小さい値を用いるのが妥当であると考えられるため、牌譜データを用いて $N_{\text{dahai}}/N_{\text{dahai-max}}$ の推定を行う。すなわち、牌譜の中でその局にアガリをできなかったあるプレイヤのプレイヤ節点に注目し、その節点での $N_{\text{dahai-max}}$ とそのプレイヤが実際に打牌できた回数の比を教師データに用いる。その際、特徴量を $\phi_{\text{dahai-ratio}} = \phi_{\text{ryuukyoku}}$ とし、確率の推定は自分の打牌回数で場合分けする。通常のロジスティック回帰分析と異なり、教師信号が $0, 1$ の整数ではなく 0 から 1 の実数となるが、パラメータの学習は通常のロジスティック回帰分析と同じように行うことが可能である。学習により得られたパラメータを $w_{\text{dahai-ratio}}$ と書くと、探索に用いる打牌回数は

$$N_{\text{dahai}} = \lceil N_{\text{dahai-max}} \sigma(w_{\text{dahai-ratio}} \cdot \phi_{\text{dahai-ratio}}) \rceil \quad (23)$$

となる。

利得 $U_{\text{not-agari}}$ は以下のように求める。はじめに、1 人麻雀の根節点における降り方策 1 人麻雀での放銃確率を $p_{\text{ori-houjuu}}$ とする。この時、降り方策 1 人麻雀における牌種それぞれの放銃確率は全て 1 人麻雀の根節点において推定される。そして、特徴

$$\phi_{\text{houjuu-after}} = \begin{pmatrix} 1 \\ \text{logit}(p_{\text{ori-houjuu}}) \\ \text{自分のフーロ回数} \end{pmatrix} \quad (24)$$

を用いたロジスティック回帰分析によって、現在の局が自分の放銃によって終了する確率を推定し、その値を $p_{\text{houjuu-after}}$ とする。これを用いて

$$U_{\text{not-agari}} = \sum_{i=2}^{4} p^{i}_{\text{agari-other}} \times \left(p_{\text{ha}} R^{i1}_{\text{agari}} + (1 - p_{\text{ha}}) \sum_{j=2}^{4} p^{ij}_{\text{agari-from}} R^{ij}_{\text{agari}} \right) \quad (25)$$

と計算する。ここで、$p_{\text{houjuu-after}}$ を p_{ha} と略記した。

アガリ方策 1 人麻雀におけるアガリ出来ない終端節点の利得を $U_{\text{not-agari}}$、打牌回数を N_{dahai} として期待値最大化方策に基づいて探索した際に得られた根節点の行動 a のアガリ確率と期待値をそれぞれ、p^{a}_{search}、E^{a}_{search} とする。この時、アガリをした場合の平均利得 $U^{a}_{\text{agari-average}}$ は

$$U^{a}_{\text{agari-average}} = \frac{E^{a}_{\text{search}} - (1 - p^{a}_{\text{search}}) U_{\text{not-agari}}}{p^{a}_{\text{search}}} \quad (26)$$

と求めることができる。

DAG の探索から求めたアガリ確率 p^{a}_{search} を元に、実際のアガリ確率 p^{a}_{agari} を推定する際にロジスティック回帰を用いる。すなわち、実際のアガリ確率推定に用いる特徴量を

$$\phi^{a}_{\text{agari}} = \begin{pmatrix} 1 \\ \text{logit}(p^{a}_{\text{search}}) \\ \text{リーチしている他家の数} \\ 1 - \prod_{i=2}^{4} (1 - p^{i}_{\text{tenpai}}) \end{pmatrix} \quad (27)$$

とする。テンパイ確率に関する積は、リーチしていない他家に対してのみとるものとし、確率の推定はそのプレイヤの打牌回数に関して場合分けを行う。

流局時のテンパイ確率 $p^{a}_{\text{ryuukyoku-tenpai}}$ も、アガリ確率と同様に推定する。すなわち、N_{dahai} 回の打牌でテンパイ確率最大方策に基づいて探索して得られた行動 a のテンパイ確率を $p^{a}_{\text{tenpai-search}}$ とし、流局時テンパイ確率 $p^{a}_{\text{ryuukyoku-tenpai}}$ 推定に用いる特徴量を

$$\phi^{a}_{\text{ryuukyoku-tenpai}} = \begin{pmatrix} 1 \\ \text{logit}(p^{a}_{\text{tenpai-search}}) \\ \text{リーチしている他家の数} \end{pmatrix} \quad (28)$$

とする。

同様に、事後放銃確率 $p^{a}_{\text{houjuu-after}}$ も推定する。これは、行動 a における打牌自体は放銃にならない場合に、その後自分の放銃によって局が終了する確率である。用いる特徴

量は、式 (24) と同じである。ただし、$p_{\text{ori-houjuu}}$ は行動 a における打牌が放銃にならかった手牌からの降り方策 1 人麻雀での放銃確率 $p^a_{\text{ori-houjuu}}$ に置き換える。

最終的に、行動 a を行い放銃しなかった場合の期待値は

$$\begin{aligned} & E^a_{\text{no-houjuu}} \\ &= p^a_{\text{ag}} U^a_{\text{aav}} + (1-p^a_{\text{ag}}) p_{\text{ry}} \left(p^a_{\text{rt}} U_{\text{rt}} + (1-p^a_{\text{rt}}) U_{\text{rn}}\right) \\ &+ (1-p^a_{\text{ag}})(1-p_{\text{ry}}) \left(p^a_{\text{ha}} U_{\text{hav}} + (1-p^a_{\text{ha}}) U_{\text{anr}}\right) \quad (29) \end{aligned}$$

となる。ただし、ag、aav、ry、rt、rn、ha、hav、anr はそれぞれ agari、agari-average、ryuukyoku、ryuukyoku-tenpai、ryuukyoku-noten、houjuu-after、houjuu-average、agari-no-relation を表す。最終的に、行動 a の期待利得 E^a は

$$\begin{aligned} E^a &= p_{\text{houjuu}}(h^a) U_{\text{houjuu}}(h^a) \\ &+ (1-p_{\text{houjuu}}(h^a)) E^a_{\text{no-houjuu}} \quad (30) \end{aligned}$$

となり、この値が最大となる選択を行う。

6. 実験結果とまとめ

この章では、提案した麻雀 AI の対局実験結果を説明する。対局では 1 人麻雀を有向非巡回グラフで表す際の牌の交換枚数は [15]、手牌のシャンテン数に基づいて決定し、0,1,2,3 シャンテンのそれぞれで 2,3,3,4 とした。ただし七対子に関する交換枚数はシャンテン数+1 枚を上限とした。4 シャンテン以上で降りを行わない場合は孤立した牌で、最も不要なものを捨てる方策を取った。候補となるのは、手持ち枚数が 1 枚の字牌、1,2,8,9 の数牌である。この中で場に見えている同種の牌が何枚あるか、その牌種がドラであるか、字牌は刻子にした時に役が付くか、数牌は同じ色の 4,5,6, が手牌に含まれているかの項目から、著者の経験則に基づいて優先順位を付けた。

対局のルールは floodgate for mahjong と同じものを用いた *3。対局相手は manue という著者らが知る範囲内ではオープンソース化されている麻雀 AI の中で最も強いものを用いた *4。対局は manue を 3 体用意して席順をランダムに決めて行った。対局の結果を表 2 に示す。平均順位については 2.193 ± 0.026 となり有意に強い結果が得られた。

表 2 提案手法と manue の対局結果

1 位率	0.370 ± 0.011	平均順位	2.193 ± 0.026
2 位率	0.249 ± 0.010	アガリ確率	0.235 ± 0.005
3 位率	0.200 ± 0.009	放銃確率	0.108 ± 0.003
4 位率	0.181 ± 0.009		

本研究では、麻雀 1 局の目的に応じた抽象化と価値推定によって麻雀プレイヤを開発する手法を提案した。目的に応じた抽象化では、本研究で新たに降り方策 1 人麻雀と包括方策 1 人麻雀を提案し、期待利得最大化の方策を定式化した。また、価値推定と組み合わせて統合的な麻雀プレイヤを開発し、manue との対局で大幅に勝ち越す結果を得た。本研究で用いた機械学習の手法は、テンパイ推定の手法も含めて全て 10 未満の特徴からなる学習の組み合わせであり、麻雀の知識が少ない開発者でも開発が容易である。また、麻雀のローカルルールに関する特徴が少ないことから、国際ルールの麻雀への適用も行いやすいと考えられる。

*3 水上直紀, 亀甲博貴, 万代悠作, 横山秀. floodgate for mahjong (仮). http://www.logos.t.u-tokyo.ac.jp/mjlog/ 2017

*4 Hiroshi Ichikawa https://github.com/gimite/mjai-manue

参考文献

[1] Murray Campbell, A. Joseph Hoane Jr., Feng-hsiung Hsu. Deep Blue Artificial Intelligence 134(1-2), pp. 57-83, 2002.

[2] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, et. al. Mastering the game of Go with deep neural networks and tree search. Nature 529, pp. 484-489, 2016.

[3] G. Tesauro. Programming backgammon using self-teaching neural nets, Artifcial Intelligence, 134 (1-2), pp. 181-199, 2002.

[4] Matej Moravčík, Martin Schmid, Neil Burch, et. al.DeepStack: Expert-level artificial intelligence in heads-up no-limit poker. Science 356, pp. 508-513, 2017.

[5] とつげき東北. 科学する麻雀. 講談社現代新書. 2011.

[6] 栗田萌, 保木邦仁. 1 人麻雀の有向非巡回グラフを用いた近似表現. 情報処理学会研究報告. GI, Vol. 2017-GI-37, No. 14, pp. 1-8, 2017.

[7] 栗田萌, 保木邦仁. 麻雀における他家の手牌と待ちの予測に基づく放銃確率推定. 情報処理学会研究報告. GI, Vol. 2017-GI-38, No. 5, pp. 1-8, 2017.

[8] 水上直紀, 中張遼太郎, 浦晃, 三輪誠, 鶴岡慶雅, 近山隆. 多人数性を分割した教師付き学習による四人麻雀プログラムの実現. 情報処理学会論文誌, Vol. 55, No. 11, pp. 1-11, 2014.

[9] 小松智希, 成澤和志, 篠原歩. 役を構成するゲームに対する効率的な行動決定アルゴリズムの提案. 情報処理学会研究報告. GI, Vol. 2012-GI-28, No. 8, pp. 1-8, 2012.

[10] 吉村健志, 宝珍輝尚, 野宮浩揮. タブーサーチを用いた麻雀における最適行動の探索第 21 回ゲームプログラミングワークショップ, pp. 73-80, 2016.

[11] 水上直紀, 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.

[12] Naoki Mizukami and Yoshimasa Tsuruoka. Building a Computer Mahjong Player Based on Monte Carlo Simulation and Opponent Models. Proceedings of the 2015 IEEE Conference on Computational Intelligence and Games, pp.275-283, 2015.

[13] 我妻敦, 田中哲朗. 半教師あり学習による麻雀の降り局面の判別. 第 20 回ゲームプログラミングワークショップ, pp. 143-147, 2015.

[14] 水上直紀, 鶴岡慶雅. 強化学習を用いた効率的な和了を行う麻雀プレイヤ. 第 21 回ゲームプログラミングワークショップ, pp. 81-88, 2016.

[15] 栗田萌, 保木邦仁. 有向非巡回グラフで表現された 1 人麻雀の探索アルゴリズム. 第 22 回ゲームプログラミングワークショップ, 2017 (発表予定).