Vol.2017-GI-37 No.14
2017/3/7

1人麻雀の有向非巡回グラフを用いた近似表現

栗田 萌[,a)] 保木 邦仁[1,b)]

概要： ヒューリスティック探索は人工知能 (AI) の基礎をなす技術として知られている。探索は 4 人不完全情報ゲームの麻雀 AI でも有効な手法と考えられるが、ゲーム木が大きすぎるため、これを正確に探索することは困難である。そこで、他人の行動をほとんど無視した 1 人麻雀について考えるが、それでも探索木の大きさは現実的に計算することができないほどである。本研究では、一人麻雀の探索木の節点のいくつかを同一視して、節点数が大幅に削減された有向非巡回グラフ (DAG) を探索する手法を提案する。

An Approximate Representation of Single-Player Mahjong by using Directed Acyclic Graph

Moyuru Kurita[,a)] Kunihito Hoki[1,b)]

Abstract: Heuristic searches are known as fundamental technology of artificial intelligence (AI). While search methods are expected useful in AI of mahjong, which is a four-player game with imperfect-information, the game tree is too large to search precisely. Even when we consider single-player mahjong, where actions of the other players are almost neglected, the size of the tree is still too large to search in practice. In this research, we propose a search method of a directed acyclic graph (DAG) to cut down the number of nodes in the game tree of single-player mahjong by neglecting distinctions between nodes in the tree.

1. はじめに

ゲームは勝敗が明確に定まるため、古くから AI の研究の対象とされてきた。中でも多くの完全情報ゲームにおいて、AI は人間のトッププレイヤと互角以上に戦う強さを持つ。

偶然の要素が含まれないゲームに関しては、チェスにおいて 1990 年代にディープ・ブルーが世界チャンピオンを破り、よりゲーム木の大きい将棋についても 2015 年に情報処理学会はトッププロ棋士に AI が勝ち越す可能性が高いとして、コンピュータ将棋プロジェクトの終了を宣言した。これらの背景には $\alpha-\beta$ 法によるゲーム木の探索や、駒の位置関係から評価関数を算出するための機械学習手法 [1] などの発展があった。また、これよりさらにゲーム木が大きくゲームの性質も大きく異なる囲碁においても、モンテカルロ木探索 [2] や深層学習 [3] を用いて 2016 年に Alpha Go が人間のトッププロに勝ち越すに至った。偶然の要素を含むゲームにおいてもバックギャモンでは 2000 年代に人間のトッププレイヤと互角以上に戦う強さになったと考えられていて、強化学習の成功例として知られている。

一方で不完全情報ゲームは、完全情報ゲームほど研究が進んでいない。例えば、状態数が比較的少ない 2 人テキサス・ホールデムは 2014 年に CFR+によって近似解が得られたが [4]、情報集合数がより多い 3 人以上のテキサス・ホールデムには適用が困難であると考えられている。

このようなゲームにおいては、プレイヤの行動や情報を抽象化することでゲームの大きさを削減する手法が提案されている [5]。この手法は、はじめにゲームの抽象化を行い、抽象化されたゲーム木における最適行動を求め、それを何らかの手法で抽象化する前の行動へ写像する手法であり、ポーカー等で一定の成功を収めている。

本研究が対象とする麻雀も、通常 4 人で行う多人数不完全情報ゲームであり、ナッシュ均衡解を求めることが困難と考えられているばかりか、他プレイヤの行動をほとんど無視して手作りだけを行う 1 人麻雀においても、最適な行

1 電気通信大学
a) mkmjai1@gmail.com
b) k.hoki@uec.ac.jp

動を求めるには至らない。そこで本研究では麻雀の AI を作成することを最終的な目的としつつ、まずは 1 人麻雀よりも探索空間が小さい、さらに抽象化された麻雀を提案する。

本論文は以下の構成になっている。初めに 2 章で麻雀のルールと用語、3 章で 1 人麻雀のルールと 1 局の結果から最終順位を予測する手法に関する関連研究を述べる。4 章では、麻雀をさらに抽象化し、元の問題の木より小さい DAG を用いて 1 人麻雀のモデルを作る具体的な手法を提案する。5 章でこの手法を用いた打牌の選択と、上級プレイヤの選択の一致率を示し、6 章でまとめと今後の展望について述べる。

2. 麻雀のルールと用語

この章では、麻雀のルールと用語について解説する。麻雀は初めに一定数の牌が配られ、そこからルールで定められた方法で牌を 1 枚ずつ交換し、**役**と呼ばれる特定の構成を作り**和了 (アガリ)** し、役に応じた点数を得るゲームである。アガリの点数は直前の手牌と最後に入手した牌とその入手方法によって定められていて、牌山から 1 枚牌を引いた場合を**ツモアガリ**、他プレイヤ（**他家**）が捨てた牌を利用する場合を**ロンアガリ**という。牌を交換する方法としては、牌山から 1 枚牌を引く（**ツモ**）場合と、他プレイヤ（他家）が捨てた牌を入手（**フーロ**）する方法がある。交換の場合は牌を入手した後不要な牌を 1 枚場に捨てる。また、特定の条件下においてプレイヤは**リーチ**を行うことができ、これは 1000 点を支払う必要があるが役になる。この 1000 点はその局でアガリした人が得る。牌山が無くなるまで誰もアガリしなかった場合を**流局**といい、各プレイヤの最後の手牌の形に応じた点数を各プレイヤが得る。リーチしたプレイヤがいる場合、支払われた 1000 点は次の局でアガリしたプレイヤの点数となる。

以上を一つの**局**として、複数の局を行いルールが定める最後の局を終了したときの点数の多さに応じて順位が決まる。一般的なルールでは開始時の各プレイヤの点数は 25000 点からはじまり、局は**場**と**局数**で分類される。**東風戦**では東 1 局から東 4 局までを、**東南戦**では東 1 局から南 4 局までを行い、最終局が終了した時に 30000 点を超えているプレイヤがいない場合追加で局が行われる。東風戦、東南戦でそれぞれ**南入**、**西入**という。この場合誰かが 30000 点を超えた時点で終了となるが、誰も 30000 点を超えずに南 4 局、西 4 局が終局した場合も終了とする。

最終局が終了した時の順位に応じて、各プレイヤに**順位点**が配分され、プレイヤの最終的な利得は順位点と点数に応じて決まる。順位点をゲームの最終結果と捉える対戦ルールも広く普及していることから、本研究では順位点をゲームの利得とする。

以下に手牌に関する麻雀の用語を解説する。手牌で 3 つ同じ牌がある場合と、連続した牌が 3 つある場合にこれを**メンツ**と呼ぶ。また同じ牌が 2 枚の場合**ヘッド**と呼ぶ。手牌がメンツ 4 つとヘッド 1 つになった場合これを**アガリ形**といい、あと 1 枚でアガリ形になる手牌を**テンパイ**という。通常テンパイは役がある場合をさすが、役が無い場合**形式テンパイ**という。通常のアガリ形ではないアガリも存在し、あと 1 枚でそのアガリになる場合もテンパイというが、その説明は省く。テンパイになるまでに必要な牌交換回数の最小値を手牌の**シャンテン数**という。

3. 先行研究

麻雀 AI の研究において、最初はどのように手作りを行うかを問題とする場合が多い。1 人麻雀の解き方としても多くの手法が提案されている。水上らの研究では、上級プレイヤの選択を教師信号としたパーセプトロンの手法を用いて、その選択に近づける方法を提案している [6]。この手法では多クラスロジスティック回帰分析を用いて期待最終順位を求める手法と組み合わせて、自分がアガリをした時の点数がほぼ確定した状態に対して適切な押し引き基準を獲得し、平均的な人間プレイヤ以上の実力を持った AI の作成に成功している [7]。また、小松ら [8] や梅津ら [9] はモンテカルロ法を用いて、フーロは無しの 1 人麻雀の手作りを行う手法を提案した。さらに論文に掲載されていないものでも、1 人麻雀を解くツールは複数存在していてインターネット上でダウンロード可能となっている [10], [11]。これらは定式化の詳細は公開されておらず、またフーロが存在する場合での期待値計算の定式化例は著者の知る限りでは存在しない。

一方でポーカーなどの不完全情報ゲームにおいては、ゲームを抽象化しゲーム木のサイズを小さくすることで最適解を求めやすくした問題を解いて、それを抽象化する前のゲームにおける行動と対応させる方法が提案されていて、大きな効果を上げている。

3.1 順位点の予測

麻雀は 1 局が終わると次の局に移行し、ルールが定める最後の局が終わった段階で最終的な順位が確定する。ある局が終わった時点での点棒状況から最終順位についての確率を求める多クラスロジスティック回帰を用いたモデルは水上らの研究 [7] によって AI に利用されている。

局の終わり方により各プレイヤの点数移動が決まり、次の局が開始するときの各プレイヤの点数が決まる。この手法を用いると、点数移動後の各プレイヤの点数から最終局後の最終的な順位を予測することができ、さらに順位点の期待値も評価することができる。

順位予想は、あるプレイヤの順位が $r \in \{1, 2, 3, 4\}$ になる確率をソフトマックス関数

$$p_r(w,\phi) = \frac{e^{w_r \phi}}{\sum_{r'=1}^{4} e^{w_{r'} \phi}} \quad (1)$$

により近似的に表す。但し、重みベクトル $w = (w_1, w_2, w_3, w_4)$ は回帰により定めるパラメタで、w_r は特徴ベクトル ϕ と同じ次元のベクトルである。特徴ベクトル ϕ は次の局が始まる時の各プレイヤの点数等により表現される。ここで、複数の特徴と順位の組からなるデータ点があるとして、i 番目のデータ点を (ϕ_i, r_i) と書く。これら N_{D} 点からなるデータによくあてはまる重み w を、最尤法により求める。最小化すべき損失関数は、

$$L(w) = -\frac{1}{N_{\mathrm{D}}} \sum_{i=1}^{N_{\mathrm{D}}} \log p_{r_i}(w, \phi_i) \quad (2)$$

とする。

本研究で考える 1 人麻雀は通常の麻雀のある 1 局に対応し、これが最終局でない場合には局後に順位点が確定しない。本節で導入された局後の順位点予測は、最終局でない局を 1 人麻雀として抽象化した場合に、1 人麻雀の利得として有用である。

3.2 フーロ、ロン無しの 1 人麻雀のルール

麻雀は通常 4 人のプレイヤで行う複数の局から成るゲームである。本節では、麻雀の 1 局を抽象化したゲームである 1 人麻雀について考える。この節ではそのルールおよびゲーム進行について説明する。本節ではフーロとロンが無く、ツモのみで進行する 1 人麻雀について説明する。以下では 13 枚の牌とリーチ棒の状態が取りうる組み合わせ全ての集合を Q と書く。ここで、リーチ棒の状態は、リーチをしているかしていないかの 2 値で表される。プレイヤには 13 枚の手牌 $q \in Q$ が与えられ、牌山の残り枚数 $t_{\max}$ が知らされる。これは 1 人麻雀では通常 17 以下の整数である。プレイヤは手牌が 13 枚の時は、牌 $h \in H$ を 1 枚牌山からツモる。H は 34 の牌種と null からなる集合であり、赤ドラ牌等は考えない。この牌は見えていない牌 h の枚数を見えていない全ての牌の枚数で割った確率でツモるものとする。手牌が 14 枚の時は、それがアガリ形でない場合牌を 1 枚捨てる行動を選択する。また、手牌が 14 枚でアガリ形である場合、牌を捨てる選択の他にアガリ宣言を行うことができ、直前の 13 枚手牌と最後のツモから一般的な麻雀のルールで定義された利得がプレイヤに与えられ、その時点で 1 人麻雀は終了となる。アガリ宣言を行わずに残り順数が無くなった場合、最後に牌を 1 枚捨てて手牌が 13 枚となった時点で 1 人麻雀は終了する。この時の利得については普及したルールは存在しないが、テンパイ形であるかによらず、点数の低いアガリよりも低く設定するのが自然である。また、プレイヤには初期の手牌以外に見えているが使えない牌が存在する場合があり、これらの牌は牌山に含まれないためツモることができないものとする。以上のゲーム進行は以下のフェーズによりまとめられる。

- **start**: はじめにプレイヤに 13 枚の牌 $q \in Q$ と牌山の残り枚数が知らされ、chance へ移行する。
- **chance**: 牌山に牌が残っている場合プレイヤは牌山から牌 $h \in H$ を 1 枚ツモり player へ移行。残っていない場合ルールに基づいた利得が与えられてゲームが終了する。
- **player**: プレイヤは牌がアガリ形の場合、アガリ宣言を行うことができて agari へ移行する。アガリ宣言を行わない場合牌を捨てて chance に戻る。また、打牌により手牌がテンパイ形になる場合リーチを行うことができて、以後アガリ以外ツモ切りしかできなくなるが、アガリをした場合の利得と流局時の利得がルールに基づいて変化する。
- **agari**: ルールに基づいた利得が与えられてゲームが終了する。

プレイヤは牌山を見ることができないため、次にツモる牌を知ることはできないが、各牌をツモる確率は算出可能である。

また、1 人麻雀においても自分と他プレイヤの点数は存在すると考える。そして、1 人麻雀をあたかも通常の麻雀の現局であるかのように考え、1 人麻雀が終わった時の点数に基づき後に続く通常の麻雀の順位点の期待値を求めて、これを 1 人麻雀の利得とする。したがって、局数や本場数など順位予測に必要な点数以外の値は 1 人麻雀開始時に定められているものとする。

3.3 木を用いた 1 人麻雀のモデル

本節では、1 人麻雀をプレイヤ数 1 の不確定完全情報ゲームと見做し、ゲーム進行の分岐点と終点を木の節点として表現したモデルを説明する。各節点は様々なゲームの状態に対応していて、

(1) プレイヤが意思決定することによりゲーム進行が分岐する節点
(2) プレイヤの意思決定とは無関係におきる偶然の要素により分岐する節点
(3) ゲームが終った状態に対応する節点

の 3 種に分類される。以降、1 番目に属する節点をプレイヤ節点、 2 番目に属する節点は偶然節点、 3 番目に属する節点は終端節点と呼び区別することにする。このゲーム木は偶然節点を持つため、同じ意思決定により異なるゲーム進行が発生し得る。また、プレイヤは意思決定を行う時には、これがどの節点で為されるのものなのかを知ることができる。利得は各終端節点に対して与えられている。ゲームの進行は、初めに 13 枚の牌が配られる偶然節点に始まり、以降ツモを行う偶然節点と打牌またはアガリを選択するプレイヤ節点の繰り返しとなる。

1 人麻雀においては、プレイヤ節点はプレイヤが打牌ま

たはアガリ宣言を選択する状態であり、偶然節点はプレイヤがツモを行う状態に対応する。各節点は、プレイヤから見えている牌の種類と数に関する情報を持ち、偶然節点での分岐の確率は見えている牌の種類と数のみにより決められる。

このようにして、1 人麻雀は、利得の期待値を最大にするように枝を選択する問題としてモデル化される。形式的には、プレイヤ節点 n において利得の期待値は

$$E(n) = \max_{n_c \in C(n)} E(n_c) \tag{3}$$

で与えられ、偶然節点 n においてその期待値は

$$E(n) = \sum_{n_c \in C(n)} p(n_c) E(n_c) \tag{4}$$

と与えられる。ここで、$C(n)$ は節点 n の子節点の集合であり、$p(n_c)$ は節点 n から n_c へ遷移する非ゼロの確率である。なお、アガリ宣言を行った場合と、ツモ牌が無くなり流局した終端節点 n においては、利得が関数 U を用いて

$$E(n) = U(n) \tag{5}$$

と定められているものとする。ここで、利得 $U(n)$ は局終了時の各プレイヤの点数に基づいて決まる順位点の期待値である。

この問題は、原理的には後ろ向き帰納法によって解くことができて、プレイヤ節点における最適な行動は式 (3) を最大化する行動である。しかし実際には木が非常に大きく、探索によって選択すべき枝の組を求めることは困難と考えられる。例えば、ツモと打牌だけを考えた場合でも、ツモで 30 程度、打牌で 10 程度の分岐が存在するため、300 の $t_{\max}$ 乗程の節点を持つ。また、配牌については 1000 億程度の組み合わせが存在することが知られているため [12]、1 人麻雀を解くことはこれらの積ほどあるプレイヤ節点に対して最適な行動を求めることに相当する。

4. 提案手法

本章では、1 人麻雀の探索空間を削減し、これを解きやすい問題に置き換える手法を提案する。DAG を用いた元の問題の木より小さい 1 人麻雀のモデルを考える。目標は、最適な行動を計算機で現実的に求められるところまで 1 人麻雀のモデルを抽象化することである。

4.1 麻雀のさらなる抽象化

1 人麻雀を、偶然節点において各子節点に遷移する確率が、その節点での手牌 $\mathrm{hand}(n) \in Q$、打牌を行った回数 $t(n)$、および現在の節点 n_0 のみから定まるとして抽象化する。ここで、$\mathrm{hand}(n)$ は節点 n における手牌を取り出す関数とする。なお、n が打牌を行うプレイヤ節点である場合、$\mathrm{hand}(n)$ は牌をツモる直前の 13 枚の手牌を返す関数であるとする。

現在の節点（根節点）を n_0 とし、ある偶然節点で手牌が $q \in Q$、打牌回数が t の時に、牌 $h \in H$ をツモる確率を $p'(q, h, t, n_0)$ と書く。また、打牌を行う節点 n において直前のツモ牌を返す関数を $\mathrm{in}(n)$ とする。この関数は偶然節点 n においては $\mathrm{in}(n) = \mathrm{null}$ とする。また期待値の計算を定式化するため、節点のフェーズが打牌を行うプレイヤ節点か、打牌後の偶然節点か、アガリを行った終端節点かを表す関数を $\mathrm{phase}(n) \in \{\mathrm{player}, \mathrm{chance}, \mathrm{agari}\}$ とする。なお、流局した節点 n は偶然節点ではないが $t(n) = t_{\max}$ かつ $\mathrm{phase}(n) = \mathrm{chance}$ とする。

利得の期待値は q, h, a, t, n_0 の関数として書く。ここで、$h \in H$ はツモ牌であり、a は節点のフェーズを表す 3 値変数とする。もし、$a = \mathrm{agari}$ ならば

$$E'(q, h, \mathrm{agari}, t, n_0) = U'_a(q, h, n_0) \tag{6}$$

となる。ここで、$U'_a(q, h, n_0)$ は現在の点数状況、アガリを行う直前の手牌 q、最後にツモった牌 h のみに依存する利得である。

また、残りのツモが無くなり最後の打牌を行って流局となった場合、すなわち $a = \mathrm{chance}$ かつ $t = t_{\max}$ かつ $h = \mathrm{null}$ であるとき

$$E'(q, \mathrm{null}, \mathrm{chance}, t_{\max}, n_0) = U'_r(q, n_0) \tag{7}$$

となる。ただし、$U'_r(q, n_0)$ は最後の手牌と現在の点数状況だけに依存する利得である。これは、テンパイしているかどうかで決まる流局時の利得を表す。

牌をツモしたプレイヤ節点、すなわち $a = \mathrm{player}$ かつ $h \neq \mathrm{null}$ では

$$\begin{aligned} &E'(q, h, \mathrm{player}, t, n_0) \\ &= \max_{(q_c, h_c, a_c) \in C'(q, h, \mathrm{player})} E'(q_c, h_c, a_c, t+1, n_0) \end{aligned} \tag{8}$$

となる。ここで、$C'(q, h, \mathrm{player})$ は、手牌 q において牌 h をツモった後に行うことができる打牌またはアガリ宣言を行った後の全ての手牌 q_c、ツモ牌 h_c、フェイズ a_c の組からなる集合である。アガリを行った場合打牌の回数は増えないが、利得の値に影響しないため形式的に子節点は全て $t+1$ とした。また、アガリ宣言を行った場合には q_c は q、h_c は h、a_c は agari になる。アガリ宣言を行わなければ q_c は打牌後の手牌、h_c は null、a_c は chance になる。

最後にツモを行う直前の節点、すなわち $a = \mathrm{chance}$ かつ $t \neq t_{\max}$ かつ $h = \mathrm{null}$ であるとき

$$\begin{aligned} &E'(q, \mathrm{null}, \mathrm{chance}, t, n_0) \\ &= \sum_{h_c \in H} p'(q, h_c, t, n_0) E'(q, h_c, \mathrm{player}, t, n_0) \end{aligned} \tag{9}$$

となる。

これらの抽象化は、もし元の 1 人麻雀のプレイヤ節点 n_{p}、

アガリを行った終端節点 n_{a}、偶然節点 n_c 全てに対して

$$p(n_{\mathrm{p}}) = p'(\mathrm{hand}(n_{\mathrm{p}}), \mathrm{in}(n_{\mathrm{p}}), t(n_{\mathrm{p}}), n_0) \quad (10)$$

$$U_{\mathrm{a}}(n_{\mathrm{a}}) = U'_{\mathrm{a}}(\mathrm{hand}(n_{\mathrm{a}}), \mathrm{in}(n_{\mathrm{a}}), n_0) \quad (11)$$

$$C'(\mathrm{hand}(n_p), \mathrm{in}(n_p), \mathrm{player})$$
$$= \{(\mathrm{hand}(n_c), \mathrm{in}(n_c), a(n_c)) | n_c \in C(n_p)\} \quad (12)$$

$$H = \{\mathrm{in}(n) | n \in C(n_c)\} \quad (13)$$

を満たすのであれば、木の節点 n に対して

$$E(n) = E'(\mathrm{hand}(n), \mathrm{in}(n), \mathrm{phase}(n), t(n), n_0) \quad (14)$$

である。しかし、実際にはツモの確率は現在の手牌だけでなくいままでの捨て牌にも依存し、またアガリの利得も一発や海底を考慮すると条件 (10)、(11) は満たされない。本節の抽象化により得られる $E'(q, h, a, t, n_0)$ は、利得の期待値を与えない。しかし、もし $p(n)$ や $U_{\mathrm{a}}(n)$ を良く近似する $p'(q, h, t, n_0)$ や $U'_{\mathrm{a}}(q, h, n_0)$ を設計することができたならば $E'(\mathrm{hand}(n), \mathrm{in}(n), \mathrm{phase}(n), t(n), n_0)$ が $E(n)$ の良い近似になると期待できる。

この近似の範囲内で、プレイヤ節点 n_0 では、

$$E'(\mathrm{hand}(n_c), \mathrm{in}(n_c), \mathrm{phase}(n_c), t(n_c), n_0) \quad (15)$$

を最大化する $n_c \in C(n_0)$ を選択する。

4.2 拡張 1 人麻雀

本節では、1 人麻雀よりも通常の麻雀にルールが近い、フーロとロンの効果を含んだ拡張 1 人麻雀を定義する。拡張 1 人麻雀も通常の麻雀の 1 局を抽象化したものである。プレイヤの他に、他家が 1 人存在しツモ切りだけを行うものとする。プレイヤはポン、チー、ロンを行うことが可能である。簡単のため、カンについては考えない。今までの 1 人麻雀と異なり、偶然節点はツモを行う節点と、相手が打牌をする節点の 2 種類が存在する。またプレイヤ節点も、打牌とアガリの選択を行う節点とフーロ、ロンの選択を行う節点が存在する。ゲーム進行は以下のようにまとめられる。

- **start**: プレイヤに 13 枚の手牌 $q \in Q$ が与えられ、$\mathrm{chance_F}$ へ移行する。
- **chance**$_{\mathrm{F}}$: 牌山に牌が残っている場合他家が牌を牌山から 1 枚場に出し $\mathrm{player_F}$ へ移行する。残っていない場合、手牌に応じた利得が与えられてゲームが終了する。
- **player**$_{\mathrm{F}}$: プレイヤは手牌と他家の捨て牌に応じてフーロやロンが可能で、ロンした場合 $\mathrm{agari_R}$ へ移行する。フーロしない場合 $\mathrm{chance_T}$ に進み、フーロした場合 $\mathrm{player_D}$ へ移行する。
- **chance**$_{\mathrm{T}}$: 牌山に牌が残っている場合プレイヤは牌山から牌を 1 枚ツモり $\mathrm{player_D}$ へ移行する。残っていない場合手牌に応じた利得が与えられてゲームが終了する。
- **player**$_{\mathrm{D}}$: プレイヤは可能なアガリまたは打牌の一つを選択し、アガリを宣言した場合には $\mathrm{agari_T}$ へ移行し、アガリをしない場合プレイヤは牌を 1 枚捨てて $\mathrm{chance_F}$ に戻る。
- **agari**$_{\mathrm{T}}$, **agari**$_{\mathrm{R}}$: ルールに基づいた利得が与えられてゲームが終了する。

ゲームの進行は手牌が配れられる偶然節点から始まり、他家がツモ切りをする偶然節点、フーロの選択を行うプレイヤ節点が続く。フーロしない場合はツモの偶然節点、打牌を行うプレイヤ節点を経て、他家がツモ切りをする偶然節点に戻る。フーロした場合ツモの偶然節点を飛ばして、打牌を行うプレイヤ節点が続いて、他家がツモ切りを行う偶然節点に戻る。フーロ無しの場合と同様に、プレイヤは牌山を見ることができないため次にツモる牌や捨てられる牌を知ることはできないが、牌種の確率分布は算出可能である。

拡張 1 人麻雀の探索木をさらに削減した DAG を探索することを考える。フーロ無しの時には利得の期待値は q, h, a, t, n_0 の関数によって表されたが、ここでは以下のような変更を行う。手牌はフーロをした手牌により構成されることもあり得る。ここで、$\hat{Q}$ はフーロまで考慮した場合の手牌全てからなる集合である。フーロを行った直後の節点 n において $\mathrm{hand}(n) \in \hat{Q}$ は、フーロを行う直前の手牌とする。

次に $\mathrm{in}(n)$ は、フーロ無しの場合ツモ牌 h のみを表していたが、フーロ直後の節点ではそのフーロ $f \in F$ を、フーロの選択を行う節点ではフーロできる牌 $h \in H$ を表し、ロンアガリの節点ではロンの牌を返す関数であるとする。ここで、F は麻雀におけるフーロの集合である。

節点 n のフェーズを表す $\mathrm{phase}(n)$ の値域は $\{\mathrm{player_D}, \mathrm{player_F}, \mathrm{chance_T}, \mathrm{chance_F}, \mathrm{agari_T}, \mathrm{agari_R}\}$ に拡張される。これらはそれぞれ、打牌またはアガリの選択を行うプレイヤ節点、フーロとロンの選択を行うプレイヤ節点、ツモを行う偶然節点、他家がツモ切りを行う偶然節点、ツモアガリを行った終端節点、ロンアガリを行った終端節点である。

t については、フーロ無しの場合と同様にプレイヤが打牌を行った回数を表すものとする。従って、ゲームの流局終了条件は牌山が無くなった時ではなく、プレイヤが $t_{\max}$ 回打牌を行った時と抽象化される。

次に期待値の計算を定式化する。アガリをした場合、すなわち $a \in \{\mathrm{agari_T}, \mathrm{agari_R}\}$ ならば

$$E''(q, h, a, t, n_0) = U''_a(q, h, a, n_0) \quad (16)$$

となる。ここで、$U''_a(q, h, a, n_0)$ は現在の点数状況、アガリを行う直前の手牌 q、アガリ牌 h、アガリがツモかロン

かを表す変数 a のみに依存する利得である。

ツモまたはフーロの後のプレイヤ節点、すなわち $a = \mathrm{player_D}$ であるとき

$$E''(q, i, \mathrm{player_D}, t, n_0) = \max_{(q_c, i_c, a_c) \in C''(q, i, \mathrm{player_D})} E''(q_c, i_c, a_c, t+1, n_0) \quad (17)$$

となる。ここで、$i \in H \cup F$ はこの節点の直前のツモまたはフーロを表す変数である。また、$C''(q, i, \mathrm{player_D})$ は、手牌 q においてツモまたはフーロ i を行った後に行うことができる打牌またはアガリ宣言を行った後の全ての手牌 q_c、ツモまたはフーロ i_c、フェイズ a_c の組からなる集合である。

フーロ、ロンの選択を行う節点、すなわち $a = \mathrm{player_F}$ であるとき

$$E''(q, h, \mathrm{player_F}, t, n_0) = \max_{(q_c, h_c, n_c) \in C''(q, h, \mathrm{player_F})} E''(q_c, h_c, a_c, t, n_0) \quad (18)$$

となる。ここで、$C''(q, h, \mathrm{player_F})$ は、手牌 q において他家が牌 h を打牌した後に行うことができるフーロ、ロンを行った後、またはフーロもロンも行わなかった全ての場合における手牌 q_c、フーロ h_c、フェイズ a_c の組からなる集合である。

また偶然節点については、ツモ前の節点と他家打牌前の節点が存在し、それぞれ

$$E''(q, \mathrm{null}, \mathrm{chance_T}, t, n_0) = \sum_{h \in H} p''_T(q, h, t, n_0) E''(q, h, \mathrm{player_D}, t, n_0)$$

$$E''(q, \mathrm{null}, \mathrm{chance_F}, t, n_0) = \sum_{h \in H} p''_F(q, h, t, n_0) E''(q, h, \mathrm{player_F}, t, n_0) \quad (19)$$

と書ける。

拡張 1 人麻雀の近似の範囲内で、プレイヤ節点 n_0 では、

$$E''(\mathrm{hand}(n_c), \mathrm{in}(n_c), \mathrm{phase}(n_c), t(n_c), n_0) \quad (20)$$

を最大化する $n_c \in C(n_0)$ を選択する。

4.3 探索空間の制限

4.1 と 4.2 節では麻雀の 1 局を抽象化した DAG を示したが、これでもグラフのサイズが大きすぎるため探索は困難である。実際 13 枚の手牌にはフーロが無い場合でも 1000 億程度の組み合わせが存在し [12]、節点の数はこれの $t_{\max}$ 倍程度存在する。そこで、根節点 n_0 において行うべき行動を決定するために探索空間を制限することを考える。制限は以下の 2 種の行動回数に課せられる。

(1) 変換回数 m_{c}: 変換回数とは、牌をツモして別の牌を捨てる行動と、フーロして牌を捨てる行動の合計回数を言う。根節点からの変換回数が大きすぎる節点も実現確率は低いと考えられる。なお、最初に牌が余っている状態から牌を捨てる行動は変換回数には含めないものとする。

(2) フーロ回数 m_{f}: フーロ回数についても、必要な牌のほとんどをフーロで手に入れる確率は低いため、制限を付けても打牌の選択に大きな影響は出にくいと考えられる。このフーロ回数は、打牌を考える根節点 n_0 から探索する節点までに行うフーロの回数であって、n_0 の段階で既に行っているフーロについては含めないものとする。

アガリに向かう場合に、初期の手牌のシャンテン数よりあまりに大きいシャンテン数を持つ節点を経由するのは現実的ではない。そこで、ある手牌のシャンテン数が m_{s} である時、テンパイ形に手牌を変えるためにこの数の変換が必要となるため、探索の範囲を $m_{\mathrm{s}} + m_{\mathrm{c}} \leq m_{\mathrm{c-max}}$ を満たす節点に限定する。この $m_{\mathrm{c-max}}$ は n_0 でのシャンテン数以上の値になっている必要がある。これは、n_0 のシャンテン数未満である場合、探索の範囲にテンパイ形を含むことができず、アガリの利得を反映することができないためである。また、$m_{\mathrm{c-max}}$ と n_0 のシャンテン数の差は、探索が考慮するいわゆるシャンテン戻しの回数である。フーロについても閾値 $m_{\mathrm{f-max}}$ を設定し、これよりフーロが多い手牌の探索は行わないことにする。このように探索の節点を制限することにより、DAG の節点数を大幅に削減することが可能になる。

5. 評価実験

この節では、前節で説明した 1 人麻雀の近似探索法を利用した打牌選択と、上級プレイヤの選択の一致率を評価する。始めに終端節点の評価値の設定を多クラスロジスティック回帰を用いて行う方法を説明し、次に偶然節点における各事象の起こる確率を定める。最後に天鳳 [13] の鳳凰卓のデータを利用して、提案手法の打牌選択と牌譜の一致率を調べる。

5.1 終端節点の評価値の設定

この節では、終端節点の評価値に関して具体的な値について考える。ここでは著者が用いた 24 クラスの多クラスロジスティック回帰分析により、最終順位についての確率を求める手法を説明する。この手法を用いた確率推定機は著者が既にインターネット上で公開している [14]。

プレイヤを A,B,C,D としそれぞれ東 1 局の東家、南家、西家、北家とする。プレイヤの点数はリーチ棒を除けば合計で 100000 点になるので、特徴ベクトル ϕ を

$$\phi = \begin{pmatrix} 1.0 \\ (\mathrm{score_A} - \mathrm{score_B})/10000 \\ (\mathrm{score_A} - \mathrm{score_C})/10000 \\ (\mathrm{score_A} - \mathrm{score_D})/10000 \end{pmatrix} \quad (21)$$

により定義する。ここで score は各プレイヤの局開始時の点棒である。局については、各局ごとに独立に学習を行うため、特徴ベクトルには含めない。また、本場や供託については大きな影響は与えないと考えて上述の 4 次元の特徴量だけを用いた。機械学習の目的は、与えられたベクトルから順位を予想することであり、その順位は上から ABCD となるものから DCBA となるものまで 24 種類あるため、この結果を教師信号として用いる。すなわち、ゲーム終了時の全体の順位が $s \in \{1, \cdots, 24\}$ になる確率をソフトマックス関数

$$p'_s(w, \phi) = \frac{e^{w_s \phi}}{\sum_{s'=1}^{24} e^{w_{s'} \phi}} \tag{22}$$

により近似的に表す。重みベクトル $w = (w_1, \cdots, w_{24})$ は回帰により定めるパラメタで w_s は 4 次元のベクトルである。最小化すべき損失関数は、N_{D} 点からなるデータ点集合の i 番目のデータ点を (ϕ_i, s_i) と書くと

$$L'(w) = -\frac{1}{N_{\mathrm{D}}} \sum_{i=1}^{N_{\mathrm{D}}} \log p'_{s_i}(w, \phi_i) \tag{23}$$

である。

24 クラスの多クラスロジスティック回帰分析を用いた場合、プレイヤの順位が r となる確率は

$$\begin{aligned} p'_r(w, \phi) &= \frac{\sum_{s=1}^{24} b_{rs} e^{w_s \phi}}{\sum_{s=1}^{24} e^{w_s \phi}} \\ b_{rs} &= \begin{cases} 1 & \text{クラス } s \text{ においてプレイヤの順位が } r \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{24}$$

となる。ここで、b_{rs} は注目するプレイヤの東 1 局の席に依存する。

表 1 に南 4 局開始時における点数状況から、起家のプレイヤの順位をテストデータ数 $N_{\mathrm{D}} = 9000$ で予測した際の予測順位と結果順位の割合表を示す。なお、学習は各局に対して独立に行うのであって、この例では 9000 局の南 4 局開始時点の特徴量と結果の集合を教師データとして利用している。また、比較対象として起家のプレイヤの順位そのものを多クラスロジスティック回帰の教師信号とした 4 クラスの学習に関する結果も表 2 示す。ここで w はロジスティック回帰分析の最適化パラメータであり、最急降下法を 100 ステップ行った後に、ニュートン法を 8 ステップ行い最適化した。

双方を比較すると 24 クラスに拡張したことによって精度が向上していることが確認できる。クラス数 24 の学習では起家の順位予想に関する式 (2) の損失関数そのものではなく、式 (23) の損失関数の最小化を行ったにも関わらず、式 (24) を用いた式 (2) の値が 4 クラス学習の予測よりも小さいという結果が得られた。

表 1 24 クラスロジスティック回帰を用いた南 4 局における起家順位予想

南 4 局順位予測 クラス数 24		予測 1 位	予測 2 位	予測 3 位	予測 4 位
結果	1 位	0.196	0.030	0.010	0.005
	2 位	0.054	0.159	0.035	0.010
	3 位	0.007	0.043	0.160	0.037
	4 位	0.002	0.007	0.039	0.204
損失関数：$L = 0.695$					

表 2 4 クラスロジスティック回帰を用いた南 4 局における起家順位予想

南 4 局順位予測 クラス数 4		予測 1 位	予測 2 位	予測 3 位	予測 4 位
結果	1 位	0.177	0.056	0.007	0.000
	2 位	0.053	0.148	0.052	0.007
	3 位	0.004	0.067	0.121	0.056
	4 位	0.000	0.008	0.067	0.176
損失関数：$L = 0.835$					

拡張 1 人麻雀をさらに実際の麻雀に近づけるため、ゲームの始めに仮想的に実際の麻雀と同じ 4 人の点数状況と現在の局数と自風が設定されるものとする。また、ゲーム全体としては東南戦を想定し、順位点は floodgate for mahjong[15] のものを利用し 1 位 30、2 位 10、3 位-10、4 位-30 とした。24 クラスの多クラスロジスティック回帰を用いて、拡張 1 人麻雀の終端節点の利得を以下のように定める。

- ツモアガリ：次局の開始状況が一意的に定義できるため、その時点での順位点期待値を利得とする。
- ロンアガリ：拡張麻雀では誰からロンしたかが定まらないため、3 人からロンした場合の順位点期待値の平均を利得とする。
- 流局：自分以外が全員テンパイという想定で、次局開始時の順位点期待値を利得とする。リーチをした場合自分の点数が 1000 点低いものとして順位点期待値を計算する。

5.2 確率値の設定

本節では探索の際に必要となる確率の値について説明する。節点 n_0 において既に手牌以外で見えている牌 h の枚数を v_h、初期の手牌を q_0、手牌 q に含まれる牌 h の枚数を表す関数を $n_h(q)$ と書くことにすると、手牌が q の時の牌 h の残り枚数は $4 - v_h - \max\{n_h(q_0), n_h(q)\}$ となるため、

$$p''_T(q, h, t, n_0) = \frac{4 - v_h - \max\{n_h(q_0), n_h(q)\}}{\sum_{h' \in H} 4 - v_{h'} - \max\{n_{h'}(q_0), n_{h'}(q)\}} \tag{25}$$

とした。t に依存する確率を考えた方が実際の麻雀に近づくと考えられるが、ここでは簡単のためこの表式を用いる。フーロについては、手牌 q において牌 h がロン、またはポンできる場合 $p''_F(q, h, t, n_0) = 3p''_T(q, h, t, n_0)$ とし、チーの

み可能な場合 $p''_F(q,h,t,n_0) = p''_T(q,h,t,n_0)$ とした。この場合、式 (19) における確率の総和が 1 を超えてしまうが、通常の手牌においてフーロ可能な牌の種類は、牌全体の数に対して小さな数であるため問題にはならない。実際、式 (19) において確率の総和が 1 を超えてしまう現象は、手牌 q においてフーロ、ロンできる牌の集合を $H_{\mathrm{F},q}$ と書いて、

$$\begin{aligned} & E''(q, \mathrm{null}, \mathrm{chance}_{\mathrm{F}}, t, n_0) \\ = & \sum_{h_{\mathrm{f}} \in H_{\mathrm{F},q}} p''_F(q, h_{\mathrm{f}}, t, n_0) E''(q, h_{\mathrm{f}}, \mathrm{player}_{\mathrm{F}}, t, n_0) \\ + & \left(1 - \sum_{h_{\mathrm{f}} \in H_{\mathrm{F},q}} p''_F(q, h_{\mathrm{f}}, t, n_0)\right) \\ \times & \ E''(q, \mathrm{null}, \mathrm{chance}_{\mathrm{T}}, t, n_0) \end{aligned} \quad (26)$$

とすることで改善される。

5.3 実験結果

最後に実験の結果について述べる。提案した手法では、放銃リスクなどは全く考慮せず自分の手牌を進める選択を行うため、牌譜との一致率を調べる際に他家はリーチもフーロも行っていない局面だけを選択した。また、手牌のシャンテン数を 3 シャンテン以下のものに限定した。探索の制限条件である $m_{\mathrm{c-max}}$ は、3 シャンテンの手牌では 3 とし、シャンテン数が 3 より小さい手牌ではシャンテン数 +1 と設定した。また、フーロ回数については全ての手牌に対して $m_{\mathrm{f-max}} = 2$ とした。表 3 に順目 t ごとの打牌一致率を示す。ここでは打牌回数 t で場合分けを行い、一つの t につき 4000 のテストデータを用いた。

表 3　提案手法と上級プレイヤの打牌一致率

t	一致率	t	一致率	t	一致率	t	一致率
1	0.49	5	0.60	9	0.64	13	0.71
2	0.53	6	0.61	10	0.66	14	0.74
3	0.57	7	0.61	11	0.68	15	0.75
4	0.57	8	0.63	12	0.71	16	0.74

結果を見ると局の序盤は一致率が低いものの、打牌回数が増加するにつれて一致率も向上する傾向が見られた。序盤は選択の優劣のつきにくい孤立牌を捨てる場合が多いことが序盤の一致率の低さに影響していると考えられる。

6. おわりに

本研究では、有向非巡回グラフを用いた 1 人麻雀の定式化を行い、その近似探索法を示した。また、1 局が終了した時の順位予想を 24 クラスの多クラスロジスティック回帰分析を用いて、麻雀のルールに関する詳細なルールを導入せずに精度向上につながることを示した。1 人麻雀の計算結果を、通常の 4 人の麻雀における降りる必要が無い場合の上級プレイヤの選択と比較した結果 6 割程度の一致率を示した。

麻雀は実際には 4 人で行うことが主で、その場合相手の手が進んだ時に降りるなどの技術が必要になり、その部分についてはこの研究の範囲外として今後の課題とする。

参考文献

[1] Kunihito Hoki, and Tomoyuki Kaneko Large-Scale Optimization for Evaluation Functions with Minimax Search. Journal of Artificial Intelligence Research 49, pp. 527-568, (2014).
[2] Rémi Coulom. Efficient selectivity and backup operators in Monte-Carlo tree search. In *5th International Conference on Computers and Games*, pp. 7283, 2006.
[3] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of Go with deep neural networks and tree search. Nature 529, pp. 484-489, 2016.
[4] Oskari Tammelin. Solving Large Imperfect Information Games Using CFR+. *arXiv:1407.5042* 2014.
[5] Tuomas Sandholm. Abstraction for solving large incomplete-information games. In *AAAI'15 Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, pp. 4127-4131, 2015.
[6] 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.
[7] 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.
[8] 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.
[9] 海津純平, 成澤和志, 篠原歩. 1 人麻雀における打ち方を考慮した評価指標に関する研究. 第 20 回ゲームプログラミングワークショップ, pp. 172-178, 2015.
[10] あら. 一人麻雀練習機. http://mahjong.ara.black/ 2017.
[11] nisi5028. 一人麻雀計算機. http://epsilon69399.blog20.fc2.com/ 2017.
[12] らすかる. 麻雀の数学. http://www10.plala.or.jp/rascalhp/mjmath.htm 2017.
[13] 角田真吾. 天鳳. http://tenhou.net/ 2017.
[14] 栗田萌. 麻雀順位予想計算機. http://critter.sakura.ne.jp/ 2017.
[15] 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.