

最中限における戦略学習の試み

大室 光^{1,a)} 横山 大作^{1,b)}

概要：最中限とは 3 人プレイのトランプゲームである。ルールはシンプルだが、多くの場面で中間の順位を狙う必要がある、独特の難しさがあるゲームである。大域的な戦略をたててちょうど良い程度の得点を狙い、局所的な戦術を用いてその得点を実際に獲得していく、という組み合わせが必要となり、コンピュータプレイヤ作成には複雑な要求があるゲームであるといつてよい。本研究では、ヒューリスティックな解法によるプレイヤを作成し、ランダムプレイヤとの対戦を用いてその有効性を探る。

キーワード：最中限, ヒューリスティックプレイヤ, 機械学習

1. はじめに

最中限は不完全情報のトランプゲームである。トランプゲームは世界でも数多く存在し、その中でも不完全情報ゲームの占める割合は大きい。そして、それらのゲームの多くが、『手札を早くなくす』、『得点を多く狙う』、『強いカードを揃える』など単純な目標となっている。

それに対し、最中限の目標は『ちょうど良い程度の得点を狙う』となっており、他のゲームと質が異なる。ちょうど良い程度の得点を狙うトランプゲームは他にも存在し、ブラックジャックなどが挙げられるが、最中限では多くの場面でそれが要求されるため、複雑な戦略が必要となる。

そのゲームの複雑さから、最中限の基本的な戦術は確立されていない。その為、プレイヤを作成するにあたって、本研究ではゲームの戦略を次の 3 層に分けて考えることにした。

- ターン毎のカードの選択

- ラウンド内での局所的な戦術
- 各ラウンドでの目標点数を決める大域的な戦略

ランダムプレイヤ同士の対戦データの分析を行い、ランダムプレイヤのみが対戦すると 37.054 % の勝率が得られることを確認した。ランダムプレイの統計を基に、ゲームの勝利とラウンドの勝敗数、ラウンドの勝敗とターンの勝敗数の関係性に着目し、勝ちやすいラウンド勝敗数を目指すシンプルなプレイヤを作成した。

また、

- 最終ターンの選択に着目、3 ターン目で得点を調整することで、ラウンドに勝つ戦略
- ゲームの終盤に着目し、4・5 ラウンド目で任意の範囲の得点を狙うことで、ゲームスコアが中位になるように調整する戦略

を検討し、その効果を評価した結果を用いると、ランダムプレイヤに対して約 50 % の勝率が得られることを確認した。

加えて、ヒューリスティックプレイヤの作成までの経緯と、同じ不完全情報ゲームである麻雀の機械学習の研究 [2] を参考に、最中限への機械学習

¹ 明治大学

^{a)} ee177120@meiji.ac.jp

^{b)} dyokoyama@meiji.ac.jp

の導入法として、手札の状態を学習させ手札の価値を推定する手法を検討する。

2. 最中限のルール

プレイヤーは3人、ジョーカーを除く52枚のトランプを使用する。カードの順位は、 $\spadesuit K > \heartsuit K > \diamondsuit K > \clubsuit K > \spadesuit Q > \heartsuit Q > \dots > \diamondsuit 2 > \clubsuit 2 > \spadesuit A > \heartsuit A > \diamondsuit A > \clubsuit A$ とする。得点はカードに書かれている数字である。

- (1) トランプをシャッフルし3人に17枚ずつ裏向けに配る。
- (2) 残った1枚は使用しないので場外に伏せておく。次のゲーム時に復活させる。
- (3) 各プレイヤーは裏向けに1枚ずつ出し、そのカードを表にして順位を比較する。出たカードのうち真ん中の順位のカードを出した人が勝者となる。勝者は自分の出したカードを表のままにし、敗者は裏返して自分の前に置く。これをターンの勝負と呼ぶ。
- (4) ターンを3回繰り返したものをラウンドと呼ぶ。ラウンドでは、その3ターンで得た得点総計の順位が真ん中の人が勝者となる。同点の場合は同点者がいずれも勝者となる。勝者はそのラウンドの得点総計を得て、それらのカードをそのままにし、敗者はそのラウンドで表になっていたカードを全て裏返す。
- (5) 1ゲームでラウンドを5回繰り返し、その5ラウンドで得た得点総計の順位が真ん中の人が勝者となる。同点の場合は同点者がいずれも勝者となる。

3. 関連研究

最中限はそのゲームの複雑さから、一般的に確立された戦略はなく、題材とした研究は少ない。

副田らは、モンテカルロ法に基づくアプローチにより、ゲームの終盤、5ラウンド目の開始時点でゲームスコアが中位のプレイヤーが最も有利であるという経験則を示した[1]。本研究でも、ゲーム終盤でのゲームスコアが中位になるよう得点を調整することで、ゲームの勝率が上がることが確かめられた。

不完全情報ゲームへの機械学習の導入の考え方として、麻雀の研究を紹介する。栗田らは、麻雀1局の目的に応じてゲーム木を複数に抽象化し、探索結果を用いて局面や配牌の勝ちを推定する手法を提案した[2]。

4. ゲームの構造とランダムプレイヤーの解析

ゲームの構造を理解する為に、ランダムプレイヤー同士を100000回対戦させ、統計的な性質を調べた。

4.1 ゲームの勝率

最中限は3人プレイのゲームだが、勝者が2人、3人になることもあるので勝率は33%よりやや高くなる。ランダムプレイヤーの勝率は37.054%となった。本研究では、対戦相手二人をランダムプレイヤーに固定し、有効な戦略の作成を行う。このランダムプレイヤーの勝率37.054%を参考値とし、作成した戦略がこの数値をどこまで超えることができるかを評価する。

4.2 ラウンドとターンの関係

図2はラウンドの勝者がそのラウンドで得点を獲得したターン数の割合をグラフにしたものである。3ターン全てで得点してラウンドに勝利することは不可能であり、全て無得点でラウンドに勝利しても、そのラウンドで得られる得点は0点である。また、2ターンでの得点では勝率が悪いのでここでは無視する。ラウンドに勝つためには、3ターンのうち1ターンのみで得点を獲得することが最適であると考えられる。

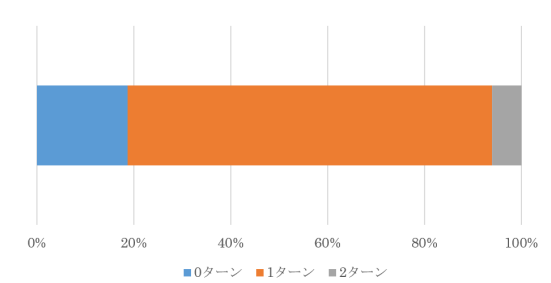


図1 ラウンド勝者の得点ターン数の割合

4.3 ゲームとラウンドの関係

図2はゲームの勝者がそのゲームで得点を獲得したラウンド数の割合をグラフにしたものである。4, 5ラウンドで得点, 全ラウンドで無得点でもゲームに勝つことは理論上可能だが, 圧倒的に少ない。ゲームに勝つためには, 5ラウンドのうち1~2ラウンドで得点を獲得することが最適であると考ええる。

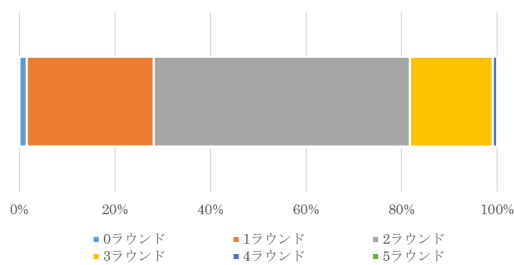


図2 ゲーム勝者の得点ラウンド数の割合

ゲームの構造は3つの層に分かれており, 本研究では各層に以下のような名前を付けて考える。

- (1) **choice**: ターンで任意の点数を取る為に最適なカードを選ぶ
- (2) **tactics**: ラウンドで任意の点数を取る為に各ターンの挙動を決める局所的な戦術
- (3) **strategy**: 各ラウンドでどの程度の点数を狙うかを決定する大域的な戦略

5. ヒューリスティックプレイヤーの構築方法

- **choice** では, 場に出すカードの条件を設定し, 手札から条件に合ったカードを選ぶ。
- **tactics** では, 目標を設定しそれに応じて各ターンで使う **choice** を選択する。
- **strategy** では, ゲームに勝つ為の仕様を設定し, それに応じて各ラウンドでどの **tactics** を選択する。

5.1 choice を決定する基本部品

まず初めにターン毎の **choice** を考える。ゲームの進行全体を見据えて考えると, 複雑な戦略を考え

る必要があるが, ターン毎の挙動に焦点を絞ると, 手札から場に出すカードを選ぶ条件を考えるだけでよい。また, **choice** で考える相手が出すカードの組み合わせは, トランプ全体から, 現時点で場に出たカードと現在のプレイヤーの手札を引いた残りのカードから考える。

- **choice0**: ランダム

主にランダムプレイヤー用の **choice**. 手札の中からランダムに選ぶ。

- **choice1**: 得点率最大

手札のカードの中で, 最も得点を得られる確率が高いカードを選ぶ。プレイヤーの手札1枚ずつに対して, 相手が出すカードの組み合わせ全てを考え, 得点を得られる組み合わせが多いカードを選ぶ。同数だった場合は, 順位が小さいカードを選ぶ。

- **choice2**: 得点率最小

手札のカードの中で, 最も得点を得られる確率が低いカードを選ぶ。考え方は上記の **choice1** と同様, 相手が出すカードの組み合わせのうちプレイヤーPが得点する組み合わせ数が最も少ないカードを選ぶ。同数だった場合は, 順位が小さいカードを選ぶ。

- **choice3**: 順位最小

手札のカードの中で, 最も小さい順位のカードを選ぶ。

- **choice4**: 順位最大

手札のカードの中で, 最も大きい順位のカードを選ぶ。

5.2 シンプルなプレイヤー

基本的な **tactics** として, “ラウンドに勝つ”と“ラウンドに負ける”の2つが考えられる。また, “ラウンドに負ける”という考え方には以下の2種類が考えられる。

(1) ラウンドの最終スコアが3人の中で最小, または最大である。

(2) ラウンドで得られたスコアが0点である。

この2つの違いは, 『最終スコアが3人の中で中位だが, 獲得した点数は0点だった』という事象を含めるか否かにある。本研究では, “ラウンドに負け

る”という tactics はラウンドで得点を取りたくない状況での使用を想定し、前述の事象を含めることにする。従って、“ラウンドに勝つ”と“ラウンドで0点を狙う”，2つの tactics を考える。

先述の通り、ラウンドに勝つ為に必要な得点回数は1回が最適である、つまり、3ターンのうちどこか1ターンで得点すればよい。ラウンドで0点を狙う為には、無得点が最大得点を狙う必要がある。また、例えばラウンドに勝ってしまったとしても、得点が0であれば条件は満たされている。従って、最大得点を狙うより無得点を狙う方に重点を置く。即ち、基本的には3ターン全てで無得点を狙い、得点してしまった場合は更に得点をして得点過多を狙えばよい。これらを踏まえ、ランダムプレイヤーの戦術 T-random, ラウンドにおいて勝ちを狙う戦術 T-take, 0点を狙う戦術 T-nottake を以下のように設定する。

- T-random

ランダムプレイヤー用の tactics, 全てのターンで choice0 を使用。

- T-take

各ターンの開始時、ラウンドスコアが0点の場合は choice1 で得点を狙い、ラウンドスコアが0点以外、即ち、既に現ラウンドで得点をしている場合は choice2 で得点を狙わない。

- T-nottake

各ターンの開始時、ラウンドスコアが0点の場合は choice2 で得点を狙わず、ラウンドスコアが0点以外、即ち、既に現ラウンドで得点をしている場合は choice1 で得点を狙う。

次にゲームに勝つ strategy を考える。先述の通り、ゲームに勝つ為に必要なラウンドの勝利回数は1〜2回で十分であることが分かっている。つまり、5ラウンドのうちどこか1・2ラウンドで得点すればよい。

これらを踏まえ、ランダムプレイヤーの戦略 S-random, ゲームに勝つ戦略 S-base を以下のように設定する。

- S-random

ランダムプレイヤー用の strategy. 全てのラウンドで T-random を使用。

- S-base

各ラウンドの開始時、ゲームスコアが0点の場合は T-take で得点を狙い、ゲームスコアが0点以外、即ち、既に現ラウンドで得点をしている場合は T-nottake で得点を狙わない。

5.3 最終ターンに着目

次に、ラウンドの最終ターンの choice に着目する。本来、3ターン目は1・2ターン目と異なり、出すべきカードの選択肢を制限できる。choice1 と choice2 だけでは3ターン目の choice が最適な選択にならない可能性がある。以下に2つの例を示す。

(1) プレイヤ P の T-take において：

3ターン目開始時点でのラウンドスコアが P が0点, A が5点, B が8点であり, P の手札の中で得点確率が最も高いカードが ♠4 である。
⇒ 得点できても最小となる

(2) プレイヤ P の T-nottake において：

3ターン目開始時点でのラウンドスコアが P が0点, A が3点, B が10点であり, P の手札の中で得点確率が最も低いカードが ♣3 である。

⇒ 得点確率に関わらず A〜2, J〜K を出せば絶対に中位にならない

つまり、3ターン目の choice は、自身のラウンドスコアだけでなく、相手のラウンドスコアとの関係性も含めたうえで考える必要があることがわかる。そこで、新たに“ターンで任意の範囲の点数を狙う”選択をする choice5 を活用し、T-take と T-nottake の精度を上げた、T-take+3t, T-nottake+3t を作成する。

- choice5：任意の得点

n1点以上 n2点以下の手札の中で、最も得点を得られる確率が高いカードを選ぶ。手札に該当するカードがない場合、どのカードを出しても条件を満たせないでランダムに返す。

- T-take+3t

1・2ターン目の開始時、ラウンドスコアが0点の場合は choice1 で得点を狙い、ラウンドスコアが0点以外、即ち、既に現ラウンドで得

点をしている場合は choice2 で得点を狙わない。3 ターン目の開始時、ラウンドスコアを中位にするために choice5 で得点調整をする。

- T-nottake+3t

1・2 ターン目の開始時、ラウンドスコアが 0 点の場合は choice2 で得点を狙わず、ラウンドスコアが 0 点以外、即ち、既に現ラウンドで得点をしている場合は choice1 で得点を狙う。3 ターン目の開始時、ラウンドスコアを最小・最大にするために choice5 で得点調整をする。

- S-base+3t

各ラウンドの開始時、ゲームスコアが 0 点の場合は T-take+3t で得点を狙い、ゲームスコアが 0 点以外、即ち、既に現ラウンドで得点をしている場合は T-nottake+3t で得点を狙わない。

ここまでの tactics と choice の関係をまとめたものが表 1 である。

表 1 tactics と choice の関係

tactics	ターン		
	1	2	3
random	choice0		
take	choice1 or choice2		
nottake			
take+3t	choice1 or choice2		choice5
nottake+3t			

5.4 最終ラウンドに着目

最後に、ゲームの終盤の tactics に着目する。ラウンドの最終ターンに着目したのと同様に、ゲームの終盤、4～5 ラウンドの開始時点で、自身のゲームスコアと相手のラウンドスコアとの関係性を考慮したうえで、ラウンドで獲得する点数を調整することを考える。

- T-takepoint

ラウンドで $n1$ 点以上 $n2$ 点以下の点数を狙う。1・2 ターン目の開始時、ラウンドスコアが 0 点の場合は choice5 で $n2$ 以下の得点を狙い、ラウンドスコアが 0 点以外の場合は choice2 で得点を狙わない。3 ターン目の開始時、ラウ

ンドスコアを中位かつ $n1$ 点以上 $n2$ 点以下にするために choice5 で得点を調整する。

- S-base+3t+4r

1～3 ラウンドの開始時、ゲームスコアが 0 点の場合は T-take+3t で得点を狙い、ゲームスコアが 0 点以外の場合は T-nottake+3t で得点を狙わない。4～5 ラウンドの開始時、ゲームスコアを中位にするために T-takepoint で得点を調整する。

ここまでの strategy と tactics の関係をまとめたものが表 2 である。

表 2 strategy と tactics の関係

strategy	ラウンド				
	1	2	3	4	5
random	T-random				
base	T-take or T-nottake				
base+3t	T-take+3t or T-nottake+3t				
base+3t+4r	T-take+3t or T-nottake+3t			T-takepoint	

6. 実験結果

6.1 tactics の評価

tactics の目的毎の性能を比較した結果が図 3 と図 4 である。“ラウンドで勝つ” tactics では、3 ターン目に得点調整を行うことでラウンドの勝率がわずかに 5 %ほど上がっていることが分かる。それに対し、“ラウンドで 0 点を狙う” tactics では、3 ターン目での得点調整の影響は少ないことが分かる。

6.2 strategy の評価

4 つの strategy の強さを比較する。各戦術に対し、ランダムプレイヤー 2 人との試合を 100000 回ずつ行なった。各 strategy の勝率をまとめたものが図 5 である。ランダムプレイヤーに対する勝率を 48.416 % 近くまでに引き上げることができた。

7. 機械学習の導入における議論

最中限への機械学習の導入法を考える。同じ不完全情報ゲームである麻雀の研究 [2] での 1 局毎の目的設定と配牌の価値推定を参考にし、strategy

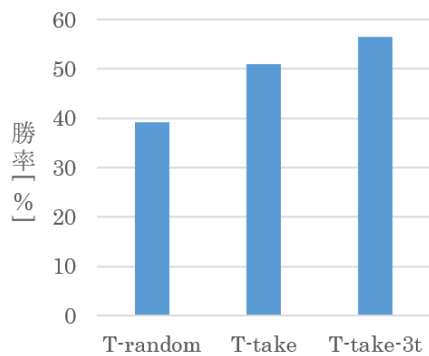


図 3 ラウンドの勝率

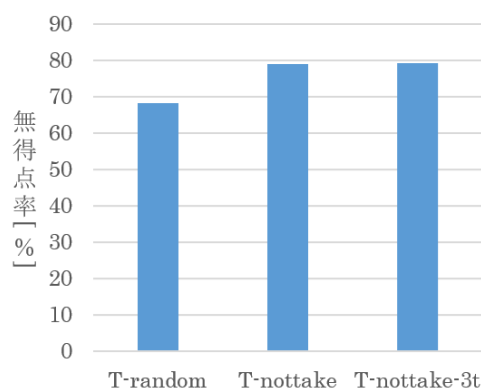


図 4 ラウンドでの 0 点成功率

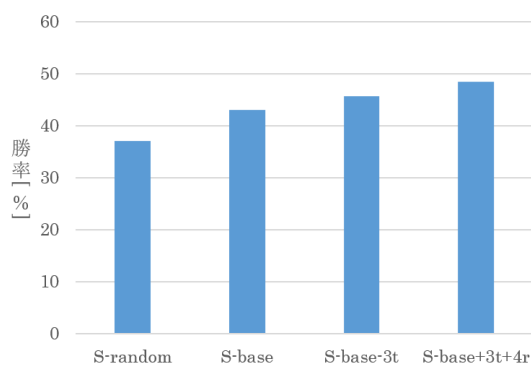


図 5 strategy 毎のゲームの勝率

と tactics で学習法を分けて考えることにした。

7.1 strategy の学習

strategy では、以下の要素から各ラウンドで取るべき最適な得点の範囲を考える。

- ゲームの目標得点
- 現在のゲームスコア
- 現在の手札の状態におけるラウンドの得点分布

ラウンドでの得点可能範囲は 0 点～13 点だが、手札の状態によって得点分布が変わる。例えば、手札に順位の大きいカードが多い場合、得点分布は大きい方に偏り、得点確率の低いカードが多い場合、得点分布は 0 に傾く。従って、ラウンド開始時点の手札の状態は、ラウンドの目標得点を決めるにあたって非常に重要な要素となる。strategy では、手札の状態毎のラウンドの得点分布を学習させ、各ラウンドでの目標点数を決める、と言う手法が最適であると考えられる。

7.2 tactics の学習

tactics では、strategy で指定された範囲の得点を取ることを考える。ここで重要となる要素は、ラウンド終了時の手札の価値である。残った手札が次のラウンドで不利になってしまうことは避けたい。ラウンド終了時点での残った手札の価値を学習させ、最適な手札の状態を求める必要がある。

どちらの学習を考えるにしても手札の状態を評価する関数が必要である。カードを評価するにあたって重要な要素として、そのカードの得点や得点確率などが挙げられる。

8. まとめ

本研究では、ランダムプレイヤ同士の対戦データの分析を行い、その統計を基にプレイヤを作成した。

実験から、“ラウンドで勝つ”仕様の tactics における 3 ターン目での得点調整、ゲームにおける 4～5 ラウンド目での得点調整が有効であることが確かめられた。

また、ヒューリスティックな解法により、ランダムプレイヤ 2 人に対して勝率 48.416 % となる戦術の作成に成功した。

最中限に機械学習を適用する手法の考察では, 麻雀の機械学習の研究を参考にし, 以下のことがわかった.

- strategy と tactics で分けて学習すべきだと考えられる.
- 手札を 1 つの状態と考え, カードを出した後の残った手札の価値を学習する, という方法が考えられる.

今後の課題として以下のことが考えられる.

- カード毎の評価値の設定法の模索.
- 手札の状態から最適な戦術を考察する.
- カードの出し方から相手の戦略を予測する.

参考文献

- [1] 副田俊介, 田中哲朗, **最中限の終盤の分析**. 情報処理学会 論文誌 GI-79, pp.31-38, 2003.
- [2] 栗田萌, 保木邦仁, **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ーの開発**. ゲームプログラミングワークショップ 2017 論文集, pp.72-79, 2017.