大富豪におけるペア温存戦略基準の獲得

○坂田　浩平 (九州工業大学大学院 情報工学府)
大橋 健 (九州工業大学大学院 情報工学研究院)

本論文では、不完全情報ゲームであるトランプゲームの大富豪を対象として、不確定な状況への適応学習について考察した。まず、予備実験により、ペア温存戦略が有効であることが確認できた。しかし、大富豪では、対戦相手・ルールによってペア温存戦略の基準が変わってくる。そこで、ペア温存戦略を動的に学習する手評価学習を実装した。手評価学習では、対戦結果に応じて、各手の評価値を更新する。対戦実験の結果、対戦相手・ルールに応じたペア温存戦略の基準が獲得出来た。

Acquisition of pair keeping strategy standard in DAIFUGOU game

*Kohei Sakata (Graduate School of Kyushu Institute of Technology),
Takeshi Ohashi (Graduate School of Kyushu Institute of Technology)

In this thesis, the adjustment study of the uncertainty to the situation was considered for the DAIFUGOU game that was the imperfect information game. At first, it was able to be confirmed that the pair keeping strategy was effective by a preliminary experiment. However, the standard of the pair keeping strategy changes in the DAIFUGOU game according to the opponent players and the rule. Then, we implemented the play evaluation learning that dynamically studied the pair keeping strategy. In the play evaluation learning, the evaluation value of each play is updated according to the game result. As a result of the experiment, the standard of the pair keeping strategy corresponding to the opponent players and the rule was able to be acquired.

1. はじめに

近年のゲームプログラミングは、ハードウェアの進歩、洗練されたアルゴリズムの開発によって、急速な発展を遂げている。チェス、将棋、囲碁のような完全情報ゲームで、コンピュータが人間に勝利する日は、刻々と近づいている。一方、麻雀やカードゲームなどの不完全情報ゲームでは、プレイヤーに隠されたゲーム状態(相手の手札) や確率的な要素(座る席、相手の出す手)が存在するなど、不確定な要素が多く、難しい課題となっている。

そこで、本論文では、不完全情報ゲームであるトランプゲームの「大富豪」を対象として、ルールや対戦相手に応じたプレー戦略をオンライン学習により獲得することを目的とする。

2. 大富豪とは？

2.1 基本ルール

大富豪は、ジョーカーを含めたカードをシャッフルして、数名の参加者に配り、手札を順番に場に出していき、いくつかの場を繰り返すことで、早く手札を無くすことを競うゲームである。カードには強さがあり、弱い順に 3、4～K、A、2、ジョーカーとなる。同じランクの数字が 2 枚以上あった場合は、同時に出すことができる(ペア)。また、同じマークの 3 枚以上の続き数字のカードも同時に出すことができる(階段)。ローカルルールと呼ばれる追加ルールの種類が多く、ルールの組み合わせや人数に応じてプレー戦略を変える必要がある。

2.2 本研究でのルール

本研究では、UECda2007[1]の Java 版開発キットを用いた。開発キットでは、サーバ側は変更できない仕様になっている。UEC 標準ルールをそのまま採用すると、あまりに複雑になってしまうので、クライアント側で、限られた手だけを生成することでルールの変更を実現した。

ただし、サーバ側の制約により、対戦人数の変更や 8 切り(8 を出すと強制的に場が流れるルール)なしとすることはできない。

本研究でのルールを以下に示す。×はサーバの制約により変更不可能なルール。○はクライアントで対応したルール。

- 対戦人数は 5 人×
- ジョーカーは 1 枚×
- ペアあり○、階段なし○、8 切りあり×
- 革命なし(よって、4 枚以上のペアはなし)○
- スペ 3 あり(スペード 3 の 1 枚で単独のジョーカーを切れる)×
- マークしばりあり×
- 全員がパスを宣言するまで（自分も含め)、場は流れない×
- 反則上がりなし×
- 階級間でのカード交換は行わない○
- ジョーカーのワイルドカード扱いなし○

3. 予備実験

まず、適応手法を考察するために、開発キットで提供されているサンプルプレイヤー(名称 Enemy) を参考に、戦略の異なるコンピュータプレイヤーを作成した。コンピュータプレイヤーの比較表を表 1 に示す。この表の「ペア温存」とは、2 枚以上あるカード崩さないという戦略を言う。「切り札考慮」とは切り札になりそうな A, 2 などの強い手札に限ってはペアを崩して 1 枚で使う戦略で、「8 切り考慮」とは、8 のペアを崩して 1 枚で使う戦略を言う。例として、PSUnder12No8 のアルゴリズムの疑似コードを図 1 に示す。表 1 の隣合うプレイヤーで 5000 試合づつ対戦させた結果、より下に記載しているプレイヤーほど強いことが分かった。ここから、「ペア温存」戦略が有効であり、切り札や 8 のペアは崩して使う戦略がより有効であることが分かった。

この予備実験を踏まえて、ペアを温存するか崩すかの基準を学習により獲得する手法を検討し、これを実装したプレイヤーを UseMEL とする。

	ペア温存	切り札考慮	8 切り考慮
Enemy	しない	しない	しない
PairSave	する(親の時しない)	しない	しない
PSInAllState	する	しない	しない
PSUnder12	する	する	しない
PSUnder12No8	する	する	する
UseMEL	学習	学習	学習

表 1．コンピュータプレイヤー比較表

```
if(自分が親){
        可能な全ての手を候補手として生成する
} else{
        場に出ているカードを考慮し、可能な手を
        候補手とする
}

ランク 11 以下の候補手の内、ランク 8 でない同一ランクにおいて、より大きなサイズのペアが可能な場合、それを候補手から外す
```

候補手の内、最低ランクの手を出す
(候補がないならパス)

図1　PSUnder12No8 擬似コード

4.　「ペア温存」戦略への適応

予備実験の結果、「ペア温存」戦略が有効であることが分かったが、相手・ルールによって「ペア温存」の基準(どの強さまでのペアを温存するのか等)が変わってくる。そこで、「ペア温存」戦略を動的に学習する「手評価学習」を考案した。

4.1　手評価学習

この学習手法は、評価値に基づいた手の選択、手の評価値更新の二つで成り立っている。

4.1.1　手の評価値

カードの各ランクにおいて、保有枚数、手(1枚出し、2枚出し、3枚出し)ごとに評価値を設定し、順次更新するようにした。更新法は 4.1.3 で示す。ランク 9 の学習結果の例を図 2 に示す。評価値は対戦経験により更新され、値が大きいほど良手であることを表す。

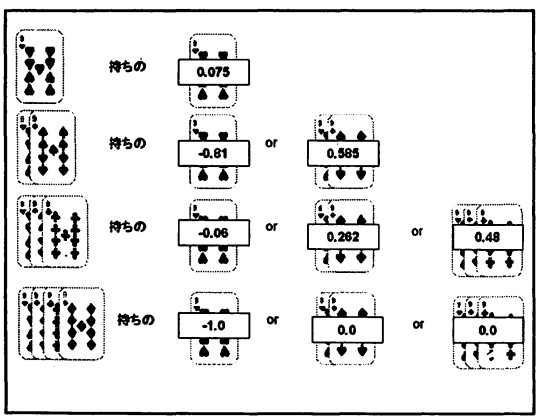

図2　ランク 9 における評価値

4.1.2　評価値に基づいた手の選択

手の選択は、同一ランクで評価値の高い手を選択するようにした。ただし、良手を発見するため、ある確率で評価値無関係に探査的な手を試すようにした。この確率を以降、ε と記述する。図 2 にランク 9 での評価値に基づいた手の選択の具体例を示す。探査的な手でない場合、図 3 の灰色の手が選択されることになる。

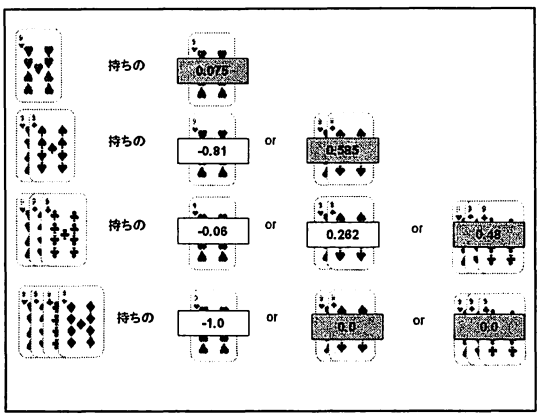

図3　ランク 9 での評価値に基づいた手の選択

4.1.3 手の評価値の更新

評価値の更新は、階級によって得られる得点を用いた（大富豪+2、富豪+1、平民±0、貧民-1、大貧民-2)。評価値更新は、以下の式(1)で行った。更新は 1 ゲームごとにそのゲーム中指した全ての手について行う。

$$\text{評価値} = \frac{\text{得点} - \text{以前の評価値}}{\text{その手が出現した回数}} \quad \text{-----(1)}$$

評価値更新の例を図 4 に示す。

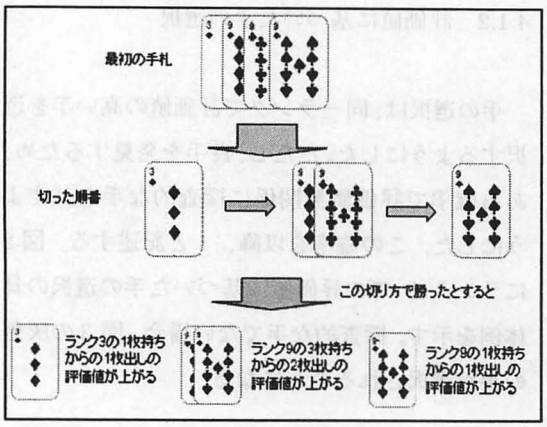

図4　評価値更新の例

4.2　テスト

前章の手評価学習を実装したプレイヤー UseMEL と自作のコンピュータプレイヤー PSInAllState と 30000 試合対戦させた。結果を図 3 に示す。また ε は 0.05 に設定した。横軸が試合数、縦軸が総得点を表す。得点は、大富豪+2、富豪+1、平民+0、貧民-1、大貧民-2 となっている。

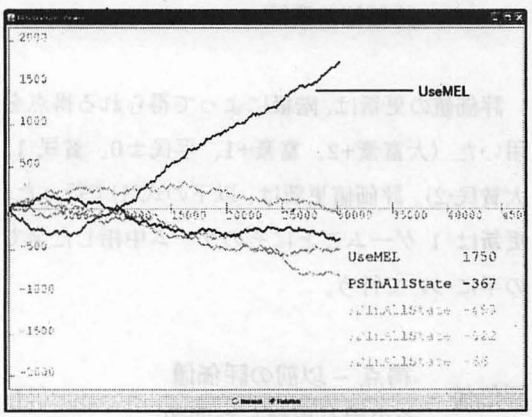

図5　UseMELvsPSInAllState

最初の負けこみが激しいが、5000 試合あたりから、グラフが上昇傾向を示し、勝ち越していることが分かる。PSInAllState に勝利できたということは「切り札考慮」、「8 切り考慮」ができたということになる。学習後のランク 3 と 2 の評価値を図 6 に示す。なお 4 枚持ちは、出現数が極端に少ないため、ここでは省略した。

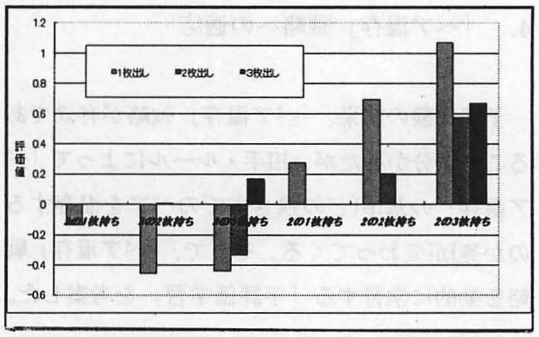

図6　学習後のランク 3 と 2 の評価値

3 は弱い手札なので、2 枚出しがある際は 2 枚出し、3 枚出しがある際は 3 枚出しが高い評価値を示しており、ペア温存する方向へ評価値が推移していることが分かる。逆に 2 は強い手札なので、2 枚出し、3 枚出しより 1 枚出しの評価値が高くなっており、崩して使う方向へ評価値が推移していることが分かる。

次に、初期状態の UseMEL と PSUnder12 を 30000 試合対戦させた。先ほどと同様で、ε は 0.05 に設定した。

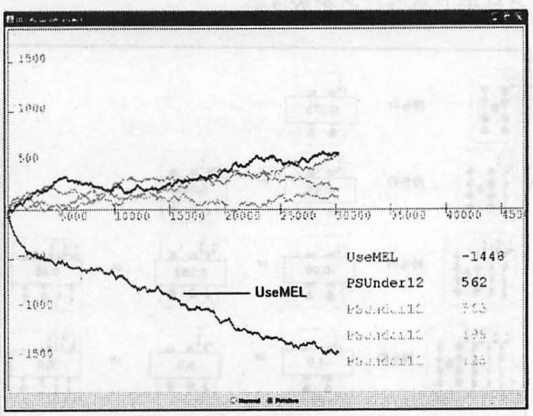

図7　UseMEL vs PSUnder12

PSUnder12 相手には、UseMEL は勝ち越すことはできなかった。しかし、グラフを見ると、最初のほうは激しく負けこみ、その後、緩やかになっていることが分かる。手評価学習では、良手を見つけるため、ε の確率で、評価値無関係に手を出す。序盤のうちは、これは必要であるが、学習が進んだ終盤では、これが敗因にもなる。そこで、学習後の評価値を用いて、探査的な手を打たない(ε =0.00 の)UseMEL と PSUnder12 を対戦させた。次ページ図 8 に結果を示す。

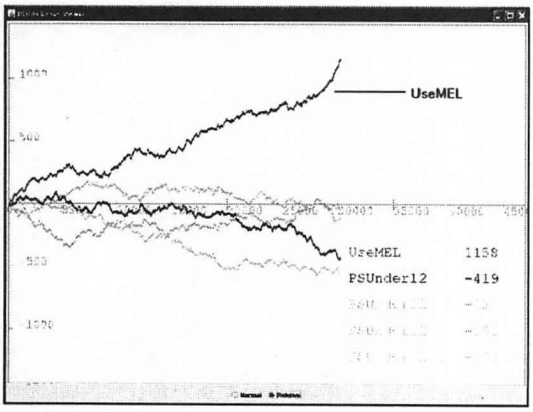

図 8　UseMEL(ε =0.00) vs PSUnder12

探査的な手を打たなければ、図 8 のように PSUnder12 にも勝ち越すことができる。PSUnder12 に勝ち越すことができたので、この評価値は 8 を上手く考慮していると考えられる。図 9 に 8 の評価値を示す。

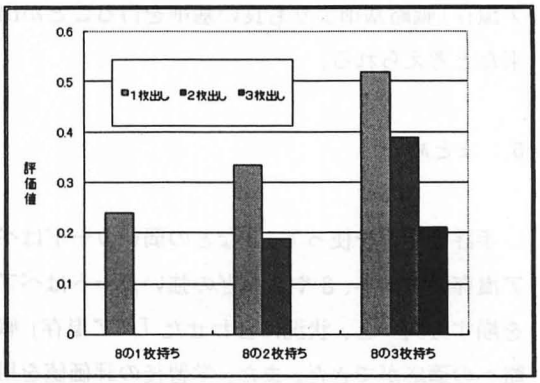

図 9　学習後のランク 8 の評価値

2 枚出し、3 枚出しより 1 枚出しの評価値が高くなっていることが分かる。これにより、親となる機会が増え、不要な手札を効率よく処理できたため、PSUnder12 に勝ち越すことができたと考えられる。

さらに、PSUnder12 と対戦して学習した UseMEL を ε =0.00 で、PSUnder12No8 を 50000 試合対戦させた。結果を図 10 に示す。

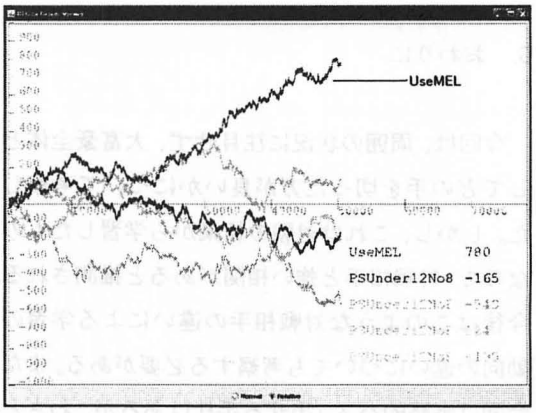

図 10　UseMEL vs PSUnder12No8

図 10 では、UseMEL は PSUnder12No8 との対戦による学習を行っていないにも関わらず勝ち越すことができた。

このことから、この UseMEL は、PSUnder12No8 のような静的に決められた「ペ

ア温存」戦略基準よりも良い基準を得ることが出来たと考えられる。

5. まとめ

手評価学習を使って、3などの弱いカードはペア温存の方向へ、8や2などの強いカードはペアを崩す方向へと、状況に合わせた「ペア温存」戦略への適応ができた。また、学習後の評価値を用いて、探査的な手を打たない状態では、現状最強のPSUnder12No8に勝てる強さを示した。自動的にこの状態にするためには、学習が進んだら探査的な手を打つ確率 ε を減らすなどの手法を実装する必要がある。

また、今回はジョーカーのワイルドカード扱いを無しにしている。これがあれば、手持ちが1枚の場合でも組み合わせてペアを作れる。対応するためには、ジョーカーを使うタイミング、ペアとして使用するかどうかなどの判断要素を盛り込む必要がある。

6. おわりに

今回は、周囲の状況に注目せず、大富豪全体としてどの手を切った方が良いかについて考察した。しかし、これは対戦の結果から学習したものなので、対戦相手と強い相関があると推測される。今後はこのような対戦相手の違いによる学習の動向の違いについても考察する必要がある。また、今回は戦略的パス（出せる手札はあるが、わざと出さない）を考慮していない。これについてもさらに検討する必要がある。

7. 参考文献

1)UECda2007:UEC(電気通信大学)が主催するコンピュータ大貧民大会
http://www.tnlab.ice.uec.ac.jp/daihinmin/2007/

2) 強化学習
Richard S.Sutton and Andrew G.Barto
三上　貞芳・皆川　雅章　共訳