大貧民ゲームの難易度調整を目的とした初期手札の強さの定量化モデル

1X15C056-0　柴田晃一
指導教員　後藤正幸

1. 研究背景・目的

大貧民ゲームとは，複数人のプレイヤーで「あがる」速さを競うポピュラーなトランプゲームである．一般的にこのゲームは人間同士で遊ぶゲームであるが，人間対コンピュータの場合，人間であるユーザがこのゲームで遊んだ場合の面白さを担保することが重要である．そのためには，ユーザのレベルに合わせて，ゲームの難易度を調整するようなハンディキャップの与え方を考える必要がある．特に，初期手札として特定プレイヤーに強いカードや弱いカードを予め配布するハンデの与え方は，ゲーム難易度への影響をコントロールできる有力な方法である．しかし，初期手札の与え方には多数の組み合わせがあり，それぞれについて採用する戦略ごとに勝率が変化していくため，どのような初期手札の与え方がどれほど勝敗に影響を及ぼすのか，定量的・客観的な評価がなされなければ適切な難易度調整は困難である．

一方，近年の人工知能技術の発展に伴い，大貧民ゲームのような多人数不完全情報ゲームにおいても，強化学習を用いて各手番における打ち手を決定するアルゴリズムの研究が進められている．このアルゴリズムを利用して，初期手札の組み合わせに対して，膨大な試合回数を試行して，各試合の勝敗を計測することで，初期手札の強さに関するデータを取得することができる．

そこで本研究では，強化学習を用いたシミュレーションに基づく初期手札の強さの定量化手法を提案する．具体的には，小沼ら [1] による TD 学習を用いたコンピュータ大貧民アルゴリズムを用いて複数のプレイヤーを用意し，初期手札によってハンデを付与したプレイヤーの勝率を計測する．これにより，様々な初期手札の勝率への影響を定量化でき，適切な難易度調整が可能となる．

2. 準備

2.1. 大貧民のルール

大貧民は，複数のプレイヤーによって「あがる」速さを競うポピュラーなトランプゲームである．一般的には,「8 切り」や「イレブンバッグ」等の特殊なルールをゲームに組み込むことで展開の意外性を高めたりするが，本研究では，最もシンプルな以下のルールを想定する．

ルール：

(I)52 枚のトランプカードを，M 人の参加者に対して，手札枚数が均等になるように配布する．最初の手番のプレイヤー (以下，親プレイヤー) が任意のカードを場に提出する．(II) 各プレイヤーが順番に,「提出ルール」に基づいてカードを場に提出する．(III) 手札が 0 枚になったプレイヤーが「あがり」，あがった順番に応じて順位が付けられる．(IV)$M-1$ 人のプレイヤーがあがった場合にゲームが終了する．

提出ルール：

(1) カードは,「3,4,5,....,K,A,2」の順番に強さが決められており，場の一番上のカードよりも強いカードであれば提出できる．(2) カードを提出しない場合，パスとなる．全員がパスをした場合，最後に場に手札を提出したプレイヤーが新たに親プレイヤーとなる．(3) 親プレイヤーは，場に複数枚同じ数字のカードを提出しても良い．その場合，以降のプレイヤーは，同じ枚数のカードを提出しなければならない．

2.2. 従来研究

従来のコンピューター大貧民ゲームにおける研究では，期待順位を目的関数として強化学習によって各手番における行動を自動で決定するモデルが提案されている．具体的には，各手番において，自身の手札及びすでに場に提出されたカード以外のカードをランダムに他のプレイヤーに割り振り強化学習を実行する環境を構築する．この環境上で,「提出ルール」の条件を満たすカードの組である合法手を選択し，行動を決定，ゲーム終了時まで各プレイヤーをルールに従いランダムに行動させ，順位に基づいた報酬を受け取る．各手番では，この報酬の平均が最大となる行動を打ち手として決定する．

小沼らは，図 1 で示す須藤ら [2] による MC 法を用いたアルゴリズムを元に，ゲーム終了時の順位だけでなくゲーム参加人数の推移から打ち手の評価を行う手法を提案している．これにより，須藤らによるものよりも勝率の高いアルゴリズムの提案に成功している．

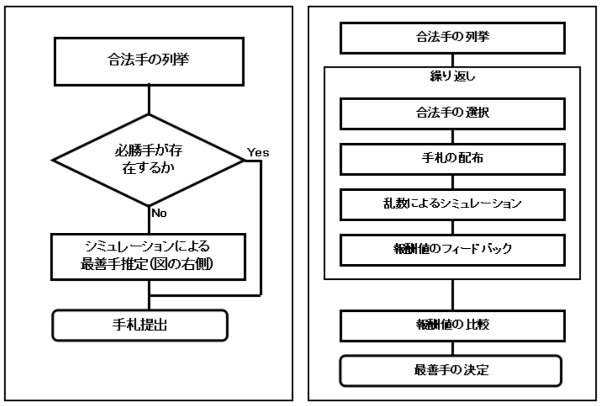

図 1: 最善手推定アルゴリズムの概略図

3. 提案手法

3.1. 概要

本研究の目的は，従来のコンピュータ大貧民ゲームの研究のように，コンピュータ同士の対戦において勝率が高くなるようなアルゴリズムの提案ではなく，強化学習を用いたシミュレーションによって，初期手札の強さの評価を通じてハンデのレベルを定量化することである．本研究では，プレイヤーに対して初期手札の一部を固定することで，ハンデを与えることを考える．そこで，同一のアルゴリズムを用いたコンピュータ同士の対戦において，配布される手札の一部を

固定したプレイヤーの勝敗を観測することで，固定した手札の強さを推定することを考える．すなわち，ランダムに配布される手札の一部を固定し勝敗を観測することで，固定した手札による勝敗への影響の客観的・定量的な評価を行う．

3.2. 提案手法のアルゴリズム

まず，同一の行動決定アルゴリズム A を持った M 人のプレイヤー，ジョーカーを除く 52 枚のトランプ x_{all} を用意する．強さを評価したい手札の組を x_{target} とし，1 ゲーム毎にプレイヤー 1 に予め配布し固定する．全ゲームプレイ回数を N，現時点のゲーム回数を $n(n \in \{1, 2, ..., N\})$ として，プレイヤー 1 の n 回目のゲームにおける順位を配列 $rank[n]$ に格納していく．以上の方針を踏まえて，提案アルゴリズムは以下のようになる．

初期手札固定ゲーム

STEP1: $n = 1, rank = 0$ として，初期化.

STEP2: x_{target} をプレイヤー 1 に対して配布し，x_{all} から x_{target} を取り除いた組を各プレイヤーの手札枚数が均等になるように配布.

STEP3: 行動決定アルゴリズム A に従って，ゲーム終了まで各プレイヤーに行動をとらせる．終了時のプレイヤー 1 の順位を $rank[n]$ に格納.

STEP4: $n = N$ であれば，配列 $rank$ を出力して終了．さもなければ，n をインクリメントし，STEP2 に戻る.

このようにして得られた $rank$ には，x_{target} が固定して配られた際の全ゲームの順位が格納されている．従って，この統計量を取ることによって，x_{target} をハンデとした場合の強さを定量的に評価することが可能となる.

4. 実験

4.1. 分析条件

提案手法のアルゴリズムを用いて，プレイヤー 1 の一部の初期カードを固定した際の勝敗を分析する．行動決定アルゴリズムは小沼らの手法を用いた．ゲーム回数は $N = 10,000$，プレイヤー人数は $M = 4$ である.

ここで，数字が i の j 枚のカードを x_i^j とする．1 枚のカードの種類による影響を調べるために $x_{target} = x_i^1(i \in \{1, 2, 3, ..., 13\})$ とした実験と，ある 1 種類のカードについてその配布枚数による影響を調べるため $x_{target} = x_3^j(i \in \{2, 3, 4\})$ とした実験の計 16 通りの実験を行う．

4.2. 結果と考察

4.2.1. カード種類による比較

まず，1 枚のカードの種類による影響を調べるために，$x_{target} = x_i^1(i \in \{1, 2, 3, ..., 13\})$ とした 13 組の平均順位について比較を行う．表1に結果を示す.

表 1: 初期カード 1 枚をハンデとした場合の平均順位

配る数字	3	4	5	6	7
平均順位	2.6974	2.6439	2.6096	2.5707	2.5395
配る数字	8	9	10	J	
平均順位	2.4759	2.4405	2.4379	2.419	
配る数字	Q	K	A	2	
平均順位	2.3863	2.4001	2.4206	2.4137	

表 1 の結果から，全体の大まかな傾向として，固定するカードの強さが強くなるにつれ平均順位が上がっていくことがわかる．また，固定するカードの強さが 1 段階変化することによる平均順位の変化量は，カードが弱いほど大きくなる傾向にある．さらに，今回の結果で平均順位が最も良くなったのは，固定するカードが最も強い 2 の場合ではなく Q の時である．これは，ゲームの試行回数が不十分なことによる誤差が生じたことも考えられる．しかし，この結果は，絵札以上の強いカードでは，それらの差異は微妙であることを示している.

4.2.2. カード枚数による比較

次に，ある 1 種類のカードについてその配布枚数による影響を調べるために，$x_{target} = x_3^j(j \in \{1, 2, 3, 4\})$ とした 4 組の平均順位について比較を行う．表 2 に結果を示す.

表 2: 初期手札に 3 を配布した場合の平均順位

配る枚数	1	2	3	4
平均順位	2.6974	2.8458	2.8964	2.9882

表 2 を見ると，固定する数字が 3 のカードの枚数が増えるに従い平均順位が悪化することがわかる．これは，固定する枚数が増えることで，他プレイヤーに表 1 で平均順位が最も悪くなることが示された 3 が配られにくくなり，他プレイヤーの平均順位が良くなっているためと考えられる．また，枚数の変化による平均順位の変化量は，固定する枚数が増えるに従い小さくなる．すなわち，弱いカードを配布することによる平均順位への影響は，1,2 枚配布するときは大きいが，3,4 枚をまとめて配布しても，それ以上に著しく弱くなる訳ではないことがわかる.

5. 考察

4 章の実験結果は，実際の大貧民ゲームにおいて初期手札によるハンディキャップを作成する際に重要な示唆を与える．例えば，コンピュータとの対戦で自身の初期手札を弱く設定しようとした場合，直感的には，3 の手札を 4 枚配ることが考えられるが，4.2.2 節の結果から，3 を 4 枚配ると手札をまとめて出す機会が増えるため，戦略によってはゲームが簡単になってしまうという意図しない結果を引き起こす可能性がある．この場合，自身の初期手札を弱くし，ゲームの難易度を上げるためには，3,4,5 といった比較的弱めのカードを重複しないように配ることが適当であると言える.

6. まとめと今後の課題

本研究では，大貧民ゲームにおける初期手札のハンデが勝率に与える影響を定量化するためのモデルとして，強化学習の導入したシミュレーションに基づく手法を提案した．今後の課題として，人間同士の対戦を正確に再現するためのより人間らしいプレイを行う行動決定アルゴリズムの提案が挙げられる.

参考文献

[1] 小沼 啓, 本多 武尊, 保木 邦仁, 西野 哲朗, “コンピュータ大貧民に対する差分学習法の応用”, 研究報告ゲーム情報学（GI）,2012-GI-27,2012.

[2] 須藤 郁弥, 篠原 歩,“モンテカルロ法を用いたコンピュータ大貧民の思考ルーチン設計”，第 1 回 UEC コンピュータ大貧民シンポジウム,2009.