

# 大貧民における偶然手番感度

西野 順二<sup>1,a)</sup> 西野 哲朗<sup>1,b)</sup>

**概要:** 大貧民プレイヤープログラムなど不完全情報ゲームでは他プレイヤーの手の情報を推定することが重要と考えられるが、局面によっては推定の必用が無い場合もある。本稿では不完全情報ゲームの性質を計る指標として偶然手番感度を提案し、これを用いて多人数不完全情報ゲームの大貧民の特性を調べることを目的とする。偶然手番感度は、相手手札の可能性を決める偶然手番がある場面での着手の利得に与える影響度であり、情報集合に含まれる局面可能性による利得の振れ幅を数値化したものである。大貧民を縮小した単貧民について、2人から5人に2から5枚を配布し合計12枚の全ての配布パターンについてその完全探索結果から、相手手札の可能性に対して最適着手は変わらず、偶然手番感度が低いことを示した。53枚5人プレイヤーにおいても、モンテカルロシミュレーションによる推定利得から偶然手番感度を求める実験によってやはり偶然手番感度が低いことを示した。

## Chance Node Sensitivity of the game of Computer Daihinmin

### 1. はじめに

二人完全情報ゲームとその人工プレイヤーに関する研究は、1997年のDeepBlueによるチェス以来[1]、計算機の能力向上も相まって近年の将棋、囲碁においてブレイクスルーとも言える急激な進化をみた。その一方で、多人数ゲームや不完全情報ゲームについては知見も少なくまだ緒のたばかりである。

多岐に渡る多人数不完全情報ゲームの特性を計る一つの指標として偶然手番感度を提案した[2]。本稿では、偶然手番感度を多人数不完全情報ゲームであるカードゲーム大貧民に適用して、そのゲームの特性を検討する。2006年から行われている不完全情報多人数ゲームのコンペティションであるコンピュータ大貧民大会[3]、[4]のクライアントを対象とする。

不完全情報ゲームでは、未知情報に関する知識が有用であることが多い。その有用さの度合いは個々のゲームによって異なっており、それがゲーム自体の特性となっている。とくに徐々に未知情報が明らかになるトリック型のカードゲームでは、着手決定において未知状態の推定を

用いればより有効な意思決定が可能となるように思われる[5]。しかしながら実験的には必ずしも有効と断定できない局面もあることが報告されている[6]。

一方、対象とする大貧民は、この偶然手番の影響が少なく偶然手番感度の低いゲームであることを大貧民を縮小した単貧民の分析と、コンピュータ大貧民による実験によって明らかにする。偶然手番感度が低ければ、情報集合内での状態推定をする必要は少なくなる。有限の計算時間のなかで意思決定を行うときには状態推定にかかる時間を減らせることは有益であり、このような課題への対応方法の方針を与える。

### 2. 不完全情報多人数ゲーム

本研究で対象とするカードゲーム大貧民は多人数ゲームと不完全情報ゲームの両方の特性を併せ持ち扱いの困難なゲームの一つである。ゲームの状態も大きく、たとえば53枚5人のゲームの場合、カードの配布という偶然手番で与えられる初期状態は $53!/(11!)^3(10!)^2 = 5.1 \times 10^{33}$ である。

本節では多人数ゲームおよび不完全情報ゲームについて、本論文で提案する偶然手番感度に関わる事項について説明する。

#### 2.1 多人数ゲーム

ゲーム理論では、多人数ゲームとは3人以上のプレイヤー

<sup>1</sup> 電気通信大学 情報理工学研究所 総合情報学専攻  
Dept. of Informatics, The University of Electro-Communications

a) nishinojunji@uec.ac.jp

b) nishino@uec.ac.jp

によって行われるゲームである。具体的には大貧民や麻雀、ダイヤモンドゲーム (Chinese checker, Stern-Halma) などがある。戦略型の多人数ゲームでは混合戦略により Nash 均衡解を得ることができる。一方で二人完全情報ゲームと異なり、展開型ゲーム木が完全に書けるとき均衡解は求まるが、実用的な最適着手は安定して求まらない性質がある。

### 2.1.1 部分ゲーム完全な Nash 均衡解

多人数ゲームを含む種々のゲームが与えられたときの最適な行動は、すべてのプレイヤーが相手が最適な行動をするときと仮定したとき個々の期待効用を最大化することであり均衡解で与えられる。Nash 均衡 [7] とは戦略の組のうちで、すべてのプレイヤーにとって自分の戦略だけをどのように変更しても自身の利得が増加しない状態をさす。

展開型のゲームでは、ゲーム木の先読み探索によって Nash 均衡解を得ることができる。多人数ゲームでは min-max の多人数ゲーム版である  $max^n$  探索 [8] により Nash 均衡点が求まり、全体に矛盾のない部分ゲーム完全均衡な解となる。

多人数ゲームのゲーム木は、ゲームの局面を表したノード  $N_i$  とその集合  $N = \{N_i\}$ 、そこで可能な着手  $M = \{m_j\}$ 、その着手による局面の遷移関数  $f: N \times M \mapsto N$  と、ゲーム終了にあたるノードとそのときの利得ベクトル  $G_h$  の集合からなる。具体的な木は、ゲームの初期状態を根ノード  $N_0$  とし遷移関数で順につないだものである。ゲームの進行とはこのノードを手に応じて遷移していくことであり、状態が進みゲームの決着が着いた局面を葉ノードと呼ぶ。葉ノードに到達すると  $k$  人の各プレイヤーが受け取る利得からなる利得ベクトル  $G = (g_1, g_2, \dots, g_k)$  が定まる。

$max^n$  探索は、あるプレイヤー  $x$  の手番のノード  $N_x$  において、そこで可能な  $L$  個の着手  $\{m_1, \dots, m_L\}$  のそれぞれの先のノードの利得  $\{G_1, \dots, G_L\}$  が定まっているとき、 $x$  にとっての最大利得を与える利得ベクトル  $G_i = (g_1^i, g_2^i, \dots, g_k^i)$  を与える手  $i$  を選び、 $G_i$  をそのノード  $N_x$  の利得ベクトルとする。先読み探索により全てのノードの利得を求め、各ノードの利得を与える着手全体の組が Nash 均衡解となる。このときノード  $N_x$  では式 (1) が成り立つ。

$$g_x^i \geq g_x^j, 1 \leq j \leq L \quad (1)$$

### 2.1.2 多人数ゲームのタイプブレイク問題と paranoid 法

ここでは多人数ゲームには特有のタイプブレイク問題について述べ、これを回避する  $max^n$  paranoid 探索法について紹介する。

$max^n$  探索は、各プレイヤーノードにおいてそのプレイヤー  $x$  の利得にしか着目しないため、ノード  $N_x$  の利得ベクトル  $G_x = (g_1, g_2, \dots, g_x, \dots, g_k)$  中のノードプレイヤーの利得  $g_x$  が等しい利得ベクトルを与える着手が複数存在することがある。この複数の選択肢を選ぶ基準はプレイヤー  $x$  にとっては全く存在せず、プレイヤー  $x$  の恣意的な選択によって他

のプレイヤーの利得を左右することになる。

この状態をタイプブレイクと呼び、多人数ゲームに特徴的な性質である。タイプブレイクの全ての点で代替可能な手はどれをとってもよく、全体の本を矛盾無く構築すればどれも完全均衡な Nash 均衡解となる。タイプブレイクの選び方によって、最終的に根ノードで取るべき最適手とその利得は大きく変化することがあるがその場合もそれぞれ完全均衡点であることは保証される。実際のゲームが進展してタイプブレイクのノードまできたときにどの手を選ぶかは、再びそのノードのプレイヤーの任意となり、探索時には意図しない手と結果を導く不安定性が  $max^n$  法にはある。

$max^n$  paranoid 探索法 [9] は、この不安定性を考慮した根ノードのプレイヤー  $a$  の意思決定法として、式 (1) に加え、式 (2) に従った利得の伝播を行う探索法である。

これは  $L$  個の手の中うち、 $max^n$  により式 (1) に従ってノードプレイヤー  $x$  の利得を最大化する手の利得の中から式 (2) のとおり、根ノードのプレイヤー  $a$  の利得が最小となる  $G_i = (\dots, g_x^i, \dots, g_a^i, \dots, g_k^i)$  を選んでノードの利得として探索を行う方式である。便宜上並べてあるが  $x$  と  $a$  の大小は規定しない。

$$g_a^i \leq g_a^j, 1 \leq j \leq L \quad (2)$$

すべての他のプレイヤーが必ず自分に不利な手を選択する、という paranoia 的判断による他者モデルを入れた探索である。

## 2.2 不完全情報ゲーム

不完全情報ゲームは、ある時点でプレイヤーが置かれた局面が、複数の可能性のうち一つに定まらないゲームである。このような可能性の集合を情報集合と呼ぶ。

局面の実現主観確率をもとに期待利得を求め、確率を含んだ中で Nash 均衡解を定義することができる。このモデル化をベイジアンゲームと呼び、展開型の不完全情報ゲームでは先読み探索により完全ベイジアン均衡解を得ることができる。

### 2.2.1 情報集合

不完全な情報のためプレイヤーからみて確定できない複数の局面の集合を情報集合とよぶ。

展開型ゲームでの情報集合  $A$  は、プレイヤー  $j$  から見て現在の状況が見分けのつかない  $D$  個のノードの集合  $A = \{N_1, N_2, \dots, N_D\}$  である。プレイヤーは  $A$  のいずれかの局面にいることは分かるが、厳密にどのノードであるかは分からない。たとえば、53 枚のカードを裏返しのまま 5 人に配ったゲーム開始時には、情報集合  $A_0$  はすべてのカードの可能な配布の組み合わせからなり、ノードの数は  $5.1 \times 10^{33}$  である。

展開型の不完全情報ゲームでは、確定したノード  $N_i$  に代わって情報集合  $A_j$  の遷移が、各プレイヤーから見たとき

のゲームの進行となる。一方で探索にあたっては情報集合内の実ノード  $N_i$  個々の遷移を仮定して考えていかざるをえない。

### 2.2.2 偶然手番

不完全情報ゲームの多くでは、未知の情報がサイコロやランダムなカード配布などによって生成されることがある。

偶然手番とは、サイコロのような偶然によって局面が選択される時、その分岐点をゲーム木のノードとしてとらえたものである。どのプレイヤーの手番でもない偶然によって局面遷移が起こるので偶然手番と呼ぶ。

カードゲームの初期の配布は、偶然手番である。さらに、ある程度ゲームが進んだ局面を考えたとき、未知の情報の数の分岐が依然残っており、これをあらためて偶然手番における確率付き選択としてとらえることができる。

現在の局面は、相手の状況の可能性が  $D$  種類あるとすると、そのすべてのノードからなる情報集合  $A = \{N_1, \dots, N_D\}$  であり、それぞれの実ノード  $N_i$  からの探索を行い期待利得を求めることで以下で述べるベイジアン均衡解を得ることができる。

### 2.2.3 モンテカルロサンプリング

展開型の不完全情報ゲームを実際に行い、プレイヤーが情報集合  $A$  にいるとき、最適な着手を探索によって求めることを考える。

このとき実ノード  $N_i$  を仮定することが必要である。これをモンテカルロサンプリングと呼ぶ。

ひとたびモンテカルロサンプリングしてノードを選択し、局面を仮定すればその仮定のもとで探索や利得の計算ができる。さらに、全ての可能な局面を生成してあわせ期待利得を導くことで、完全ベイジアン均衡解を求めることができる。

ただし、実際の問題ではすべてを生成することは計算時間の制限から不可能なことがあり、その場合は一部分をサンプリングした近似解となる。

### 2.2.4 ベイジアン均衡

展開型の不完全情報ゲームにおいても、先読み探索によってベイジアンゲームの期待利得をもとに完全ベイジアン均衡解を得ることができる。

プレイヤー  $x$  の手番のある情報集合  $A = \{N_1, \dots, N_D\}$  においてそれぞれの実ノードの実現可能性を主観確率  $\{p_1, \dots, p_D\}$  で定める。各実ノード  $N_i$  で手  $j$  を選んだときの利得ベクトルを  $\{G_1^j, \dots, G_D^j\}$  とするとき、情報集合  $A$  における手  $1 \leq j \leq L$  の期待利得  $G_A^j$  は、式 (3) で求められる。

$$G_A^j = \sum_{i=1}^D p_i G_i^j \quad (3)$$

情報集合  $A$  に含まれる実ノード  $N_i$  での利得は、 $\max^n$  にもとづき、式 (3) で求めた  $\{G_A^1, G_A^2, \dots, G_A^L\}$  からプレイ

ヤ  $x$  の手番であるのでその利得要素  $g_{A,x}^j$  が最大となるものを選び  $G_A^i$  となる。さらにゲーム木を展開し先読み探索を再帰的に行うことでゲーム全体の探索を行ったことになる。こうして得られた均衡解がベイジアン均衡解である。

## 2.3 コンピュータ大貧民

大貧民は、日本国内で著名なトランプを用いたカードゲームであり、手札をルールにもとづいて順に消費し、手札の無くなった順位を競う。多くの場合3名以上の多人数でおこなわれ地域ごとにルールが異なることが多い。

配られた状態で、互いの手を見ることができず、最後の二人になるまでのゲーム途中では全体の配布状態は明らかでない。このため不完全情報ゲームとなっている。ルール上、必ず誰かの手札は減るため、カード全体は単調に減少する。このため有限手数で必ず終了することが保証されている。

Long らは不完全情報ゲームを大きく二種類に分けており [10]、一つはトリック型ゲームでターンごとにカードを見せ合うことでゲームが進み徐々に情報が明らかになるゲームであり大貧民もこのタイプである。最後の二人になったあとは完全情報ゲームとなる。もう一つはポーカー型ゲームで勝敗決定までのプロセスでは明らかな情報開示のほとんどないゲームである。

2006年より電気通信大学ではコンピュータ大貧民大会 [3], [4] を開催し、プレイアルゴリズムを搭載したコンピュータプレイヤーの性能を競っている。これによって、より現実的で強い多人数不完全情報ゲームのアルゴリズム研究が急速に進んでいる。特に2009年、2010年には、須藤らによるモンテカルロ探索を使用したコンピュータプレイヤー snowl[11] が優勝し、囲碁に続きモンテカルロ探索の有効性が明らかとなった。

カードの配布から後述する標準ルールに則った試合進行を自動でおこなう対戦サーバが用意され、5つのクライアントプログラムがTCP/IPで接続しオンライン対戦をおこなう。開催時によって異なるが、おおむね1,000試合から10,000試合をした総得点で勝敗を競う。得点は上がった順に5,4,3,2,1点が与えられ、1000試合の場合ならば最大5000点、最小1000点を獲得する。

本研究では、コンピュータ大貧民大会の標準ルールに則るものとし、その内容を以下に示す。

- ジョーカー1枚を含む53枚を5人に対し席順に11枚,11枚,11枚,10枚,10枚と配布する。
- 二度目以降の試合では先の試合の勝敗に応じたカード交換を行う。大富豪と大貧民は2枚ずつ、富豪と貧民は1枚ずつ交換し平民はなにもしない。大貧民と貧民からは自動的に強い順のカードが選ばれ、大富豪と富豪はそれらを受け取った状態から任意のカードをサー

バを通じて相手に渡す。

- ダイヤの3を持つプレイヤーから試合を開始し、順にカードを場に出して行く。
- 自分の手番では、すでに場にあるカードと同じ出し方で、かつ、より強いカードを出すことができる。
- カードの強さは数字で判断し、3が最も弱く、J, Q, K, A, 2の順で2が最も強い。
- カードの出し方には、1枚~4枚の同じ強さのカードの組、もしくは3枚以上の同じマークの連番(階段)がある。
- ジョーカーはスペードの3を除くすべての1枚のカードより強い。
- スペードの3は1枚のジョーカーが場に出たときに限り、ジョーカーに勝つ。
- 直前のカードと同じマークの(組み合わせの)カードを出すとシバリとなる。シバリの間は同じマーク以外は出すことができない。
- カードが出せないときはパスができる。いったんパスすると場のカードが無くなるまでパスとなる。
- 全てのプレイヤーがパスをすると場のカードが流れてなくなり、最後のカードを出したプレイヤーの手番となる。
- 場にカードがなければ何をだしてもよい。
- 8を出すと場のカードは流れてなくなる(8切り)。
- 4枚組みまたは5枚以上の階段を出すと革命となり、その試合中はジョーカー以外の強さが全て逆転する。

### 3. 偶然手番感度

偶然手番を持つゲームでの着手決定では、情報集合内の複数のノード候補のうち実際にどのノードであるかを推測するための確率分布が重要である。これはカードゲームであれば、相手の手札状態の推定に他ならない。

ところが、ゲームの局面によっては、相手の手札によらず最適着手が決定することがある。例えば最強の手札を持っていれば相手状態に無関係に確率1で勝利することが分かる。

偶然手番感度はこのような偶然手番から引き起こされる未知情報による探索への影響を計る指針である。

#### 3.1 偶然手番感度

不完全情報ゲームには、そのゲームの構造によって起因する偶然手番の影響、すなわち各ノードの実現確率による影響の大小がある。

あるプレイヤー  $x$  の情報集合を  $A = \{N_i\}, 1 \leq i \leq D$  とし、ある手  $j$  の期待利得に対する各実ノード  $N_i$  の実現確率  $p_i$  の影響の度合いを考える。プレイヤー  $x$  からみて、手  $j$  の利得  $G_j^i = (g_{i,1}^j, \dots, g_{i,x}^j, \dots, g_{i,k}^j)$  の  $g_{i,x}^j$  がどの実ノード  $N_i$  に対しても一定であるとすれば  $p_i$  がどのようになっていても、その手から得られる期待利得  $G_A^j = \sum_{1 \leq i \leq D} p_i g_{i,x}^j$

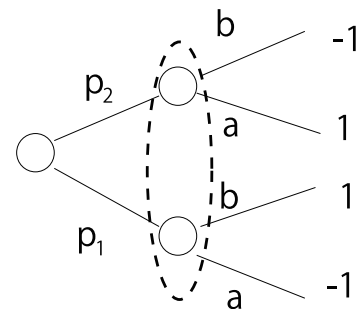


図1 偶然手番感度の高いゲーム

は一定となる。このような場合は、実現確率すなわち偶然手番の影響がないと言える。期待利得に対する偶然手番の影響は利得  $g_{i,x}^j$  の分布によって変化する。

ここで情報集合  $A = \{N_1, \dots, N_D\}$  における手  $j$  の偶然手番の影響度を式(4)によって定義する。これを偶然手番感度(Chance Node Sensitivity)と呼ぶことにする。

$$CNS(A, j) = \frac{1}{D} \sum_{i=1}^D (\bar{g}^j - g_i^j)^2 \quad (4)$$

ここで  $\bar{g}^j = \frac{1}{D} \sum_{i=1}^D g_i^j$  である。

平均からの偏差二乗の総和平均となっている。経済指標として貧富の差の状態を計るジニ係数では偏差の絶対値の総和平均をとって資産のばらつきを計っており、偶然手番感度と式の上では類似している。

$CNS(A, j)$  の値が大きいくほど感度は高くなり、 $p_i$  の変動に敏感である。 $CNS(A, j) = 0$  のときは全ての  $g_i^j$  が等しいときである。

偶然手番感度は情報集合と手を指定して定まるものであり、ゲームそのものの性質と其中でのその手の性質によって決まる。この値が高ければ、より精密な偶然手番の実現確率分布を知る必要がある。

たとえば、図1のゲームでは、点線で囲まれた二つのノードが情報集合である。 $p_1$  と  $p_2$  の値を知ることは、最適着手決定において本質的となる。 $a$  の利得は  $g(a) = -p_1 + p_2$  であり  $b$  の利得は  $g(b) = p_1 - p_2$  となるため、 $p_1 > p_2$  であるならば、 $g(a) < 0 < g(b)$  であり、逆に  $p_1 < p_2$  であるならば、 $g(a) > 0 > g(b)$  となるため、選好が逆転する。

一方、図2のゲームは手  $a, b$  ともに偶然手番感度が0であり、実現確率  $p$  の状況にかかわらず  $a$  を選択すればよい。

実際のゲームにおいては各手ごとに値が異なり、選択される期待利得の大きい(minノードにおいては小さい)手の感度が低ければ状況推定があまり必要ないと言える。

### 4. 単貧民の偶然手番感度

本研究の主目的はコンピュータ大貧民であるが、初手場面での情報集合は自分以外のプレイヤーの手の可能性で  $6.6 \times 10^{22}$  通りと大きくそのままでは計算評価ができない。このため、まず大貧民の性質を持つ計算可能な範囲の小さ

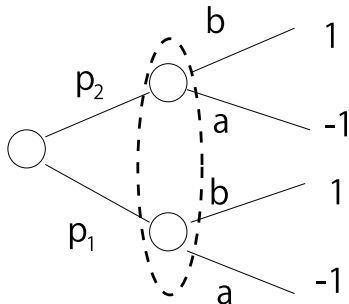


図 2 偶然手番感度の低いゲーム

表 1 全体の配布組み合わせ

枚数	2人	3人	4人	5人
1	2	6	24	120
2	6	90	2,520	113,400
3	20	1,680	369,600	$168 \times 10^6$
4	70	34,650	$63.1 \times 10^6$	$305 \times 10^9$
5	252	756,756	$11.7 \times 10^9$	$623 \times 10^{12}$

なゲームを定義し、偶然手番感度の傾向について検討する。

#### 4.1 単貧民

単貧民 [12] は大貧民の基本的なルールと構造を引き継ぎながら、分析のために縮小化したゲームである。こうした縮小モデルはそのままでは大きすぎるゲームの分析で用いられており、ポーカーの Texas Hold'em の縮小モデルである Leduc Hold'em [13] などの例がある。

使用する  $N$  枚のカードの強さは  $1, 2, \dots, N$  と全て異なるものとし、単貧民では常に単数のみの試合とする。大貧民でおこなわれる、複数カードの組による手と着手の再構成がなくなり、これによって  $N$  枚時の行動選択肢はパスを含む  $N+1$  通りに制限される。以降の分析では 12 枚までを対象とする。

本研究では相手手札については見えないものとし、そのうえで各パターンを取り出して完全情報ゲームとして探索および分析をおこなう。

#### 4.2 状態の構造

計算可能な範囲での単貧民の配布パターンを生成することを考える。

$N$  枚ずつ  $M$  人に合計  $NM$  枚を配布するとき、着目プレイヤーの手を固定すると可能な配布の組み合わせの総数は  $N$  枚ずつ  $(M-1)$  人配布とおなじく  $(M-1)!/(N!)^{M-1}$  となる。これが初期状態での情報集合の要素数であり、具体例を表 2 に示す。着目プレイヤーが受け取るカード配布の種類は  ${}_N C_N$  である。これらの配布組み合わせを 1 枚ずつ 2 人の合計 2 枚から、3 枚 4 人の合計 12 枚までを生成した。計算時間上の制約から合計 12 枚より少ないパターンのみ扱った。

#### 4.3 探索分析の結果

前述のすべての組み合わせについて、2.1.2 節で述べた  $\max^n$  paranoid 探索法によって多人数ゲーム木の完全探索をそれぞれ行い、偶然手番感度について分析を行う。

ある配布について、自手の組み合わせを一つ定める。その自手の中から可能な  $M$  個の着手について個々の番号を  $j$  とする。相手手札の  $D$  個の状態  $\{N_1, N_2, \dots, N_D\}$  のひとつひとつについて、 $j$  を初手で出したとしたときの完全探索をおこなった。ここで状態  $N_i$  を仮定し手  $j$  を出したとき得られる利得を  $g_i^j$  とする。

相手手札の状況によって、最良手は相対利得によって決められるが、それぞれで得られる実際の利得自体は異なる。このため、それぞれの状況においてそれぞれ最良手に 1, それ以外に 0 を割り当てる関数  $B: g_i^j \mapsto \{0, 1\}$  をおいて集計した。

初手  $j$  について式 (5) で与える最大の  $E(j)$  を与える手  $j$  が、全体での最良手である。

$$E(j) = \frac{1}{D} \sum_{i=1}^D B(g_i^j) \quad (5)$$

$E(j)$  が情報集合に含まれる全ての実ノード  $N_i$  で 1 であったとすると、その分散は 0 であるため偶然手番感度も 0 となりノードの実現確率  $p_i$  を求める必要なく、最良手が見つかったことになる。このとき偶然手番感度は式 (6) で与える。ただし  $\overline{B(g^j)}$  は  $B(g_i^j)$  の平均値とする。

$$CNS(A, j) = \frac{1}{D} \sum_{i=1}^D (\overline{B(g^j)} - B(g_i^j))^2 \quad (6)$$

最も感度の高い 3 枚 4 人合計 12 枚の配布での探索では、自分の手の種類が 220 通りあり、それぞれ相手の配布パターンが 1680 通りずつであった。このうち 123 通りの自分の手では、偶然手番感度がゼロであるため、半数以上で相手の配布種類を気にせず、最適着手をきめることができる。また、偶然手番感度がゼロでないパターンでも、ほとんどが 1550 通り以上の相手パターンで同じ手を採用しており偶然手番感度は 0.27 であった。最悪値は 0.45 で 1680 通りのうち 1220 通りは値が一定で偶然手番を意識する必要がなかった。残る 460 通りが他の手を単独で支持していたことになる。

以上の探索結果から、12 枚までの組み合わせで配布した状態では、3 枚 4 人、4 枚 3 人、3 枚 3 人の 3 種類以外偶然手番感度はゼロであって未知の状況の実現確率を意識せずに着手が決まることが分かった。

#### 5. 大貧民における偶然手番感度

単貧民の全探索分析で明らかになったように、大貧民型のゲームでは、偶然手番で生起する様々なノードの不確定状態に対する感度が低く、自手と残りの手全体から適当な手をきめることができると予想される。

表 2  $E(j)$  が 0 の自手パタン数 /  $n$  と偶然手番感度の最悪値/総数, \*は自明な配布, - の欄は計算量が多く未算出

枚数	2 人	3 人	4 人	5 人
1	*2/2, 0	3/3, 0	4/4, 0	5/5,0
2	*6/6, 0	15/15, 0	28/28,0	45/45,0
3	*20/20, 0	74/84, 0.36/20	123/220, 0.45/1680	-
4	*42/42, 0	420/495, 0.38/70	-	-
5	*252/252, 0	-	-	-

本節では、コンピュータ大貧民大会で優勝したプログラムをリファレンスモデルとして、偶然手番での推定を行わなくとも強いプレイヤーが構築できることを示し、実サイズの大貧民でも偶然手番感度が高くないことを検証する。また、ゲーム進展における偶然手番感度の変化についても実験的に検討を行う。

### 5.1 実験環境

対戦はコンピュータ大貧民大会と同様に 5 体のクライアントによる試合を行う。一対比較する 2 体のクライアントプログラムと、3 体のサーバ付属の default クライアントが 5 体の内訳である。評価は 1000 試合終了後の 2 つの比較プレイヤーの獲得得点  $h_1, h_2$  について (7) 式の分比を得点率  $r_{1,2}$  として採用する。

$$r_{1,2} = \frac{h_1}{h_1 + h_2} \quad (7)$$

リファレンスとして 2010 コンピュータ大貧民大会優勝クライアントの snow1 を用いた。snow1 は相手手札の推定モデルによるモンテカルロサンプリングとそこから素朴なモンテカルロ法 [14] によって最適着手を決定している。snow1 はソースプログラムが公開されており、2011 年度の大会においてもほとんどの参加クライアントがその亜種であった。また 2011 年大会でも優秀な成績を納めている。

相手手札推定モデル [11] には、ELO レーティングと囲碁のパターン学習にならった MM アルゴリズムと Bradley-Terry 推定モデル [15], [16] が用いられている。このため計算負荷が高く、core2duo 2.66GHz で 1 試合あたり 0.5 秒 (およそ一手 0.05 秒) を要する。

モンテカルロ法では式 (8) に示す UCB 値を用いてシミュレーションのコントロールを行っている。ここで  $\bar{X}_j$  は手  $j$  のプレイヤー利得の平均、 $n$  はシミュレーション総数、 $n_j$  は手  $j$  のシミュレーション回数である。平均評価が高く、シミュレーション回数の少ない手ほど UCB 値は大きくなり、選択されやすくなる。

$$UCB(j) = \bar{X}_j + c \sqrt{\frac{2 \log n}{n_j}} \quad (8)$$

### 5.2 ゲーム展開による偶然手番感度の変化

偶然手番感度のゲームの進展に対する変化について実験的に検討する。

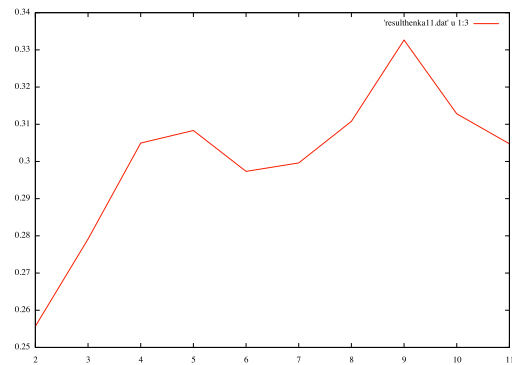


図 3 自手残枚数に対する偶然手番感度の変化

53 枚 5 プレイヤのゲーム中の各情報集合において 100 回のモンテカルロサンプリングを行い、それぞれの状況  $1 \leq s \leq 100$  での各手  $i$  に対してそれぞれ 100 回のモンテカルロシミュレーションにより推定利得  $g_i^s$  を求めた。この推定利得 100 ずつから偶然手番感度  $CNS(s, i)$  を求めた。

ゲームの進展度に対する偶然手番感度変化を図 3 と図 4 に示す。自分の手札が 11 枚から 2 枚までの変化を左図に、プレイヤー全体の合計枚数 53 枚から 5 枚までの変化を右図に示した。横軸はカードの残枚数であり右から左へゲームが進行している。

ゲームの進展にあわせ、右から左に偶然手番感度の変化を見てゆくと、全体についても自身の手についても、おおむね 0.3 程度で台状に推移し枚数の少ない終盤に向かって急激に 0 になる台形となっている。

全体 53 枚からの変化で見ると、序盤は低く中盤にむかって徐々に増加し終盤で急峻に減少する。減少への折れ点は約 16 枚であり、中盤と終盤の境界を表していると考えられる。この枚数では先上がったプレイヤーも出てきている。カード枚数の減少による可能性の低下と人数の減少による不完全性の低下の総合によって偶然手番感度が減少していることが推察される。

### 5.3 snow1 と多回数ランダムサンプリングプレイヤーの比較

多回数ランダムサンプリングプレイヤーと標準の snow1 との比較を行った。ランダムサンプリングプレイヤーは、snow1 を修正し、相手モデリングを行う部分を取り除いたものである。



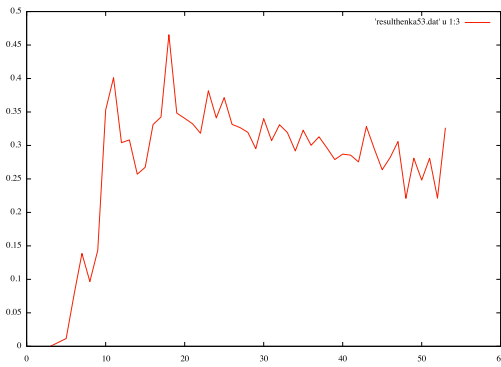


図 4 全体残枚数に対する偶然手番感度の変化

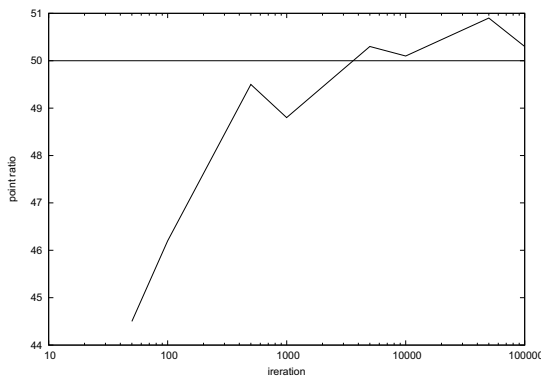


図 5 ランダムサンプリングプレイヤー得点率のプレイアウト回数に対する変化

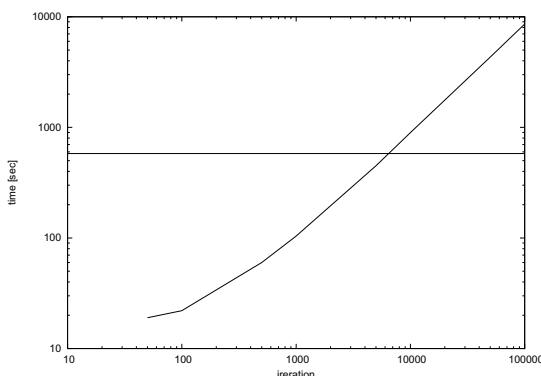


図 6 ランダムサンプリングプレイヤーの 1000 試合所要時間のプレイアウト数による変化

残カードの確率更新と、残カード確率からの相手手札推定を行わないため、snowl より高速化されており、snowl と同等の 2000 回のプレイアウトでは、1 試合あたり 0.19 秒で終了し、およそ 3 倍の速度であった。

プレイアウト回数と、対 snowl のランダムサンプリングプレイヤーの得点率の変化を図 5 に示す。

このときのランダムサンプリングプレイヤーの計算時間を図 6 に示す。ランダムサンプリングプレイヤーでは時間はプレイアウト回数に対して線形に増えている。約 7000 回で通常の snowl と同じ消費時間となった。

得点率を見ると同じ 2000 回のシミュレーションでは snowl の方が得点率が高いことから、snowl の相手手札推

定が一定の役割を役割を果たしていることが分かる。

しかし、5000 回程度のシミュレーションで 50% の得点率となり snowl に匹敵し、さらにシミュレーション回数を増やすと、安定して 50% 以上の得点率をあげて、ランダムサンプリングプレイヤーが snowl より強い。

この結果から、相手手札推定が強さにとって決定的な意味をもっているわけではないと言える。なぜなら相手手札推定機能からすると無駄と位置づけられるより幅広いノードでの探索結果は、相手手札推定によって限定された情報集合の部分集合中での最適手探索の結果と同等かそれ以上に良い手であったためである。

以上をまとめると、本実験により snowl の持つ相手手札推定機能には一定の意味があり、プレイアウト回数を削減できることがわかったがその原理は明らかではない。

一方で均質にランダムな相手手札推定であっても、回数が多くなれば snowl と同等かそれ以上強くなることもわかった。このことから、相手手札推定からはずれたサンプルノードであってもその支持する最適着手にずれが無いことが言え、実サイズのコンピュータ大貧民も偶然手番感度が十分低いことが分かった。

## 6. おわりに

本論文では、最適着手が偶然手番の実現確率にどの程度依存するかを示す偶然手番感度を定義し、この評価基準にもとづいて典型的な不完全情報多人数ゲームであるコンピュータ大貧民の特徴を分析した。

分析用に縮小した単貧民の全探索によって、小型の問題では偶然手番感度が著しく低いことを示した。

また、53 枚の実サイズ大貧民においても、ゲームの進展に対する偶然手番感度をモンテカルロサンプリングとモンテカルロシミュレーションによる推定利得から計測した。推定機構を持つ強いコンピュータ大貧民プレイヤー snowl と推定がなくプレイアウト数の多いプレイヤーの勝率比較実験では、プレイアウト数が多ければ推定と匹敵する結果を得た。これは偶然手番感度が低いことに起因するものと考えられる。

偶然手番は名目上あらゆる時点での情報集合の中の実ノードの実現確率を決定するが、実際にはこの確率はゲームが終わるか一定の進展を見せるまでは不明である。そのため、偶然手番で定まる実現確率に利得と最適着手が依存する度合いの高いゲームでは、状況の推定が重要となる。

逆に大貧民のように偶然手番感度が低いゲームでは、場面によっては状態推定にかかる労力をランダムサンプリングとモンテカルロシミュレーションの回数を増やすことに使った方が良い場合もある。実際、計算環境にもよるが snowl では、3 倍のランダムサンプリングとプレイアウトを行う方が、実 snowl で状態推定をするよりも強いことが分かった。

以上から、対象とするゲームの性質によっては、実際に推定されている相手手札等の正確さそのものではなく、モンテカルロ探索など最終的な着手決定過程も経たしきの、シミュレーション確度をあげる効果が重要であることが示された。とくに推定器の開発と調整にかけるコストには慎重になる必要がある。

不完全情報ゲームに対するUCTアルゴリズム [17], [18] も提案され、推定は行わずに状態の生成をモンテカルロサンプリングで行い、生成された木をさらにUCTでモンテカルロ探索することがおこなわれている。これは、大貧民のような偶然手番感度の低いゲームでは、有効である可能性が高いと考えられる。大貧民にかぎらず偶然手番感度という特徴量によって、UCTが利用しやすい局面かどうかの判断がつくことは開発にとって有益である。

偶然手番感度の低いゲームでは、精密な一つの正解局面の推定ではなく、その情報集合での最適な着手を与えることのできる局面の部分集合が分かればよい。ポーカーの状態縮約 [19] で提案された探索空間を削減する abstraction は、このような局面の部分集合を求めることに通じる。大貧民を含む他の多人数不完全情報ゲームでの重要な局面集合の発見方法については今後の課題である。

## 参考文献

- [1] Campbell, M., Jr., A. H. and hsiung Hsu, F.: Deep Blue, *Artificial Intelligence*, Vol. 134, No. 1-2, pp. 57 – 83 (online), DOI: 10.1016/S0004-3702(01)00129-1 (2002).
- [2] 西野順二, 西野哲朗: コンピュータ大貧民における最良手の推定について, 情報処理学会研究報告 2012-MPS-90, No. 4 (2012).
- [3] 西野哲朗: 第1回UECコンピュータ大貧民大会 (UECda-2006) の実施報告, 情報処理学会誌, Vol. 48, No. 8, pp. 884–888 (2007).
- [4] 大久保, 小林, 本多, 眞鍋, 青木, 柿下, 小松原, 西野: 第1回コンピュータ大貧民大会 (UECda-2006) の報告, 情報処理学会ゲーム情報学研究報告, Vol. GI-17, pp. 25–32 (2007).
- [5] 小田和友仁, 上原貴夫: コンピュータブリッジにおける他者のモデルを考慮したゲーム木探索の提案 (知識処理), 情報処理学会論文誌, Vol. 47, No. 11, pp. 3005–3016 (2006).
- [6] 西野順二, 西野哲朗: 大貧民における相手手札推定, 情報処理学会研究報告 2011-MPS-85, No. 9 (2011).
- [7] Nash, J.: Noncooperative games, *Ann. of Math.*, Vol. 54, pp. 286–295 (1951).
- [8] Luckhardt, C. A. and Irani, K. B.: An algorithmic solution of N-person games, *AAAI-86*, pp. 158–162 (1986).
- [9] Sturtevant, N. and Korf, R.: On Pruning Techniques for Multi-Player Games, *Proceedings AAAI-2000* (2000).
- [10] Long, J., Sturtevant, N. R., Buro, M. and Furtak, T.: Understanding the Success of Perfect Information Monte Carlo Sampling in Game Tree Search, *Proceedings of the 24th. AAAI Conf.*, AAAI, pp. 134 – 140 (2010).
- [11] 須藤郁弥, 成澤和志, 篠原 歩: UEC コンピュータ大貧民大会向けクライアント「snow1」の開発, 第2回UECコンピュータ大貧民シンポジウム講演予稿集, 電気通信大学 (2010).
- [12] 西野順二: 単貧民における多人数完全情報展開型ゲームの考察, 第12回ゲームプログラミングワークショップ, pp. 66–73 (2007).
- [13] Southey, F., Bowling, M., Larson, B., Piccione, C., Burch, N., Billings, D. and Rayner, C.: Bayes’ bluff: Opponent modelling in poker, *In Proceedings of the 21st Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 550–558 (2005).
- [14] Auer, P., Bianci, N. C. and Fischer, P.: Finite-time Analysis of the Multiarmed Bandit Problem, *Machine Learning*, Vol. 47, pp. 235–256 (2002).
- [15] Hunter, D. R.: MM algorithms for generalized Bradley-Terry models, *The Annals of Statistics*, Vol. 32, No. 1, pp. 384–406 (2004).
- [16] Coulom, R.: Computing Elo Ratings of Move Patterns in the Game of Go, *ICGA Journal*, Vol. 30, No. 4, pp. 198–208 (2007).
- [17] 三木理斗, 三輪 誠, 近山 隆: UCT 探索による不完全情報下の行動決定, 第14回ゲームプログラミングワークショップ 2009, pp. 43–50 (2009).
- [18] Sturtevant, N.: An Analysis of UCT in Multi-player Games, *Computers and Games*, Lecture Notes in Computer Science, Vol. 5131, Springer, pp. 37–49 (2008).
- [19] Zinkevich, M., Johanson, M., Bowling, M. and Piccione, C.: Regret Minimization in Games with Incomplete Information, *Neural Information Processing Systems*, Vol. 20, pp. 1729 – 1736 (2008).