Vol.2012-MPS-90 No.4
2012/9/19

# コンピュータ大貧民における最良手の推定について

西野 順二[1,a)]　西野 哲朗[1,b)]

**概要**：本論文では，不完全情報多人数ゲームであるコンピュータ大貧民における戦略決定に着目し，その最良手を求めるアルゴリズムにおいて未知の状態の推定の重要性があまり高くないことを偶然手番感度を定義し分析と実験を通じて示した．具体的には，大貧民を全探索可能なサイズに縮小した単貧民に対し，最良手の偶然手番感度がゼロないし十分小さいことを 12 枚以下のすべての配布パターンについて $max^n$ *paranoid* 探索法による全探索結果の分析から示した．さらに，コンピュータ大貧民大会で優勝したクライアントをもちい，偶然手番の状態を着手履歴から推定した場合と，単純にランダムなモンテカルロ探索回数を増やした場合の勝率比較実験を行った．この結果からコンピュータ大貧民での最良手は状態の推定によらずランダムに回数を増やしたクライアントの勝率が高いことを示し，偶然手番に対する感度が低いことを明らかにした．

NISHINO JUNJI[1,a)]　NISHINO TETSURO[1,b)]

***Abstract:*** In this article, we define a novel index, the chance node reactivity, for characterizing an imperfect information multi player game such as computer Daihinmin. An estimation of the game situation from imperfect information is not so important to select the best move for Daihinmin, where the reactivity index is quite low. We demonstrate the index feasibility by two discussions. First, we analyze on whole search result of Tanhinmin reduced Daihinmin model up to 12 cards with $max^n$*paranoid* search, and show the reactivity to situations is very low or zero. Second, we examine test matches between computer Daihinmin client with randomly Monte Carlo simulations and client with situation detector. As a result, the client with situation detector does not win so high ratio. We show that the full scale Daihinmin also has low situation reactivity.

## 1. はじめに

本論文では多数のエージェントによる不確実な情報環境下での意思決定について，最良な対応を取るときに有効な対象ゲームの性質について，偶然手番感度を提案する．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コンピュータ大貧民大会 [1] のエージェントを対象に，その最良手の探索とゲームの性質の関係について議論する．

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ゲーム [2] 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 [3]．しかしながら実験的には必ずしも有効と断定できないことが報告されている [4]．

本論文では，未知情報の源泉を偶然手番としてとらえ，その影響の度合いを偶然手番感度として定義する．偶然手番感度の高いゲームではより正確な局面の推定が有効である．

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あきらかにする．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

ゲーム理論によってエージェント間の交渉がモデル化され成果が上がっているが，その多くはチェスや将棋，囲碁

1 電気通信大学 情報理工学研究科 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

などの完全情報二人ゼロ和ゲームであった．

一方，より現実的な問題ではゲームにおける情報の不完全性と多人数の関与が必須である．これらは不完全情報ゲーム，多人数ゲームとしてモデル化できるが，二人ゼロ和ゲームと比べ，表現可能なゲームのクラスが広くゲーム単体の規模も大きくなるため，その性質は多岐に渡りまだ明らかになっていない課題も多い．

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，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 人以上のプレイヤによって行われるゲームである．具体的にはコンピュータ大貧民や麻雀，ダイヤモンドゲームなどがある．

#### 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，すべてのプレイヤが相手が最適な行動をする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Nash 均衡解で与えられる．Nash 均衡とは戦略の組のうちで，すべてのプレイヤ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 [5] 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 $k$ 人の各プレイヤが受け取る利得からなる利得ベクトル $G = (g_1, g_2, \cdots g_k)$ が定まる．

$max^n$ 探索は，あるプレーヤ $x$ の手番のノード $N_x$ において，そこで可能な $L$ 個の着手 $\{m_1, \cdots m_L\}$ のそれぞれの先のノードの利得 $\{G_1, \cdots G_L\}$ 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 $G_i = (g_1^i, g_2^i, \cdots g_k^i)$ を選び，そのノード $N_x$ の利得とする．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．

$$g_x^i \geq g_x^j,\ 1 \leq j \leq L \tag{1}$$

#### 2.1.2 多人数ゲームのタイブレーク問題と paranoid 探索法

$max^n$ 探索は，各プレイヤノードにおいてそのプレイヤ $x$ 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 $G_x = (g_1, g_2, \cdots g_x, \cdots, g_k)$ 中のノードプレイヤの利得 $g_x$ 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 $x$ にとっては全く存在せず，プレイヤ $x$ の恣意的な選択によって他のプレイヤの利得を左右することになる．

$max^n paranoid$ 探索法 [6] は，タイブレークを考慮した根ノードのプレイヤ $a$ 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) にしたがってノードプレイヤ $x$ 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 $a$ の利得が最小となる $G_i = (\cdots g_x^i, \cdots, g_a^i, \cdots g_k^i)$ 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．すべての他のプレイヤ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．タイブレークノード以下で根ノードプレイヤ $a$ の利得の下限を取ることとなり，根ノードの選択までその下限の範囲で収まることが保証される．

$$g_a^i \leq g_a^j,\ 1 \leq j \leq L \tag{2}$$

### 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．

#### 2.2.1 情報集合

不完全な情報のためプレイヤ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 $j$ 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 $A = \{N_1, N_2, \cdots N_D\}$ である．プレイヤは $A$ 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．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，各プレイヤから見たときのゲームの進行となる．

#### 2.2.2 ベイジアン均衡

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．

プレイヤ $x$ の手番のある情報集合 $A = \{N_1, \cdots N_D\}$

においてそれぞれの実ノードの実現可能性を主観確率 $\{p_1, \cdots p_D\}$ 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 $\{G^j_1, \cdots G^j_D\}$ とするとこのとき，情報集合 $A$ における手 $1 \leq j \leq L$ の期待利得 $G^j_A$ は，式 (3) で求められる．

$$G^j_A = \sum_{i=1}^{D} p_i G^j_i \tag{3}$$

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．

#### 2.2.3 偶然手番

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とき，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．

#### 2.2.4 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．このとき詳細な状況をランダムに生成して実ノード $N_i$ を仮定することが必要であり，これをモンテカルロサンプリングと呼ぶ．

### 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．

2006 年より電気通信大学ではコンピュータ大貧民大会 [1] が開催され，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．開始初年度の 2006 年はアプリオリなルールベース型のクライアントが優勝し，その後 2 年は必勝手優先の探索型が優勝した．特に 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 snowl[7] 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 試合の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 11 枚ないし 10 枚配布する．
- 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．
- ダイヤの 3 を持つプレイヤから試合を開始する．
- 各手番では場のカードと同じ形式でより強いカードを出せる．
- カードの強さは，3 が最も弱く，J，Q，K，A，2 の順で 2 が最も強い．
- 形式には，1 枚〜4 枚のカードの組，3 枚以上の連番 (階段) がある．
- ジョーカーはすべてのカードより強く, 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．
- 直前の場と同じマークのカードを出すとシバリとなる．
- 出せないときはパスができ，場のカードが無くなるまでパスとなる．
- 全てのプレイヤがパスをすると場のカードが流れる．
- 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が流れる．
- 4 枚組みまたは 5 枚以上の階段を出すと革命となり，強さが全て逆転する．

## 3. 不完全情報ゲームの偶然手番感度

不完全情報ゲームには様々な種類のゲームが含まれ，その最適な意思決定は一概に扱うことはできない．本研究では対象とする大貧民の着手決定における重要な性質として偶然手番感度を定義する．

### 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 $x$ の情報集合 $A = \{N_i\}, 1 \leq i \leq D$ で，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．たとえばプレイヤ $x$ からみて，手 $j$ の利得 $G^i_j = (g^j_{i,1}, \cdots g^j_{i,x}, \cdots g^j_{i,k})$ の $g^j_{i,x}$ 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 $g^j_A = \sum_{1 \leq i \leq D} p_i g^j_{i,x}$ 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 $g^j_{i,x}$ の分布によって変化する．

ここで情報集合 $A = \{N_1, \cdots N_D\}$ 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．これは各ノードの利得の標本分散と等しく，偶然手番感度 $R(Reactivity)$ と呼ぶことにする．

$$R(A,j) = \frac{1}{D}\sum_{i=1}^{D}(\overline{g^j} - g_i^j)^2 \qquad (4)$$

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

$R(A,j)$ の値が大きいほど感度は高くなり，$p_i$ の変動に敏感である．$R(A,j)=0$ のときは全ての $g_i^j$ が等しいときであり，偶然手番感度が低いことを表す．

これは, つぎのように考えることができる．式 (3 ) で表された情報集合 $A$ における手 $j$ の期待利得は，ノードの実現確率を並べた実現確率ベクトル $(p_1, p_2, \cdots p_D)$ と手の先のノードの期待利得を並べた期待利得ベクトル $(g_1^j, g_2^j, \cdots g_D^j)$ の内積である．ここで $R(A,j)=0$ とすると，ノードの先の期待利得ベクトルは $g^j(1,1,\cdots 1)$ であり，期待利得を表す内積は $p_i$ の分布にかかわらず $\sum p_i = 1$ の $g^j$ によるスカラー倍すなわち，$g^j$ そのものとなる．つまり，偶然手番感度が低い手 $j$ では情報集合内に多数の実ノードがあったとしても，その実現確率 $p_i$ の分布にかかわらず期待利得が一意に求まることになる．

## 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 [8] 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 [9] などの例がある．

使用する $N$ 枚のカードの強さは $1, 2, \cdots N$ 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．

### 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．着目プレイヤが受け取るカード配布の種類は ${}_{NM}C_N$ 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 人まで，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．

**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 | $11.7 \times 10^9$ | $623 \times 10^{12}$ |

### 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 $\{N_1, N_2, \cdots N_D\}$ 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．

最良手の評価を相対値とし，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 $B : g_i^j \mapsto \{0,1\}$ にもとづき集計した．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．

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

$E(j)$ 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 $p_i$ を求めずに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 $R(A,j) = \frac{1}{D}\sum_{i=1}^{D}(\overline{B(g^j)} - B(g_i^j))^2$ で与える．$\overline{B(g^j)}$ は $B(g_i^j)$ の平均値である．

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．偶然手番感度がゼロでないものも非常に小さいことが分かるった．

## 5. 大貧民クライアントによる実験

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．

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ーヤーが構築できることを示すことで，

**表 2** $E(j) = 0$ の自手パタン数 $/n$，偶然手番感度の最悪値/総数，- 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 | - | - | - |

実サイズの大貧民も偶然手番感度が高くないことを検証する．

### 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 つの比較プレイヤの獲得得点 $h_1$，$h_2$ についてその分比を得点率 $r_{1,2} = \frac{h_1}{h_1+h_2}$ として採用する．

### 5.2 snowl のアルゴリズム

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 snowl を用いた．snowl 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，優勝もできる強さをいまだ持っている．

snowl 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の単純なモンテカルロ法 [10] によって最適着手を決定している．相手手札推定モデル [7] 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 [11], [12] 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 (6) 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 $\overline{X_j}$ は手 $j$ のプレイヤ利得の平均，$n$ 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．

$$UCB(j) = \overline{X}_j + c\sqrt{\frac{2log\,n}{n_j}} \qquad (6)$$

snowl のアルゴリズムは以下の通りである．

(1) 他のプレイヤの手番で，着手から残カード確率を更新する
(2) 自手の着手決定
  (a) 合法手集合を生成する
  (b) 各手の UCB 値を初期化する
  (c) 相手プレイヤ全員の配布状態を残カード確率から生成する
  (d) UCB 値にもとづいて着手を選ぶ
  (e) 選んだ着手についてランダムシミュレーションを行い利得を得る
  (f) 得た利得から UCB 値を更新する
  (g) 規定回数 (約 2000 回) になるまで手順 (2c) に戻って繰り返す
(3) UCB 値最大の手を着手として採用する

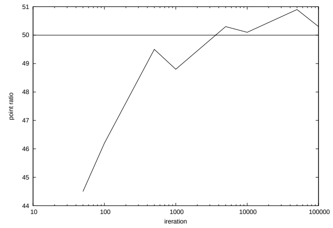


**図 1** ランダムサンプリングプレイヤの対 snowl 得点率変化

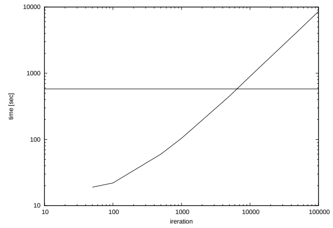


**図 2** ランダムサンプリングプレイヤの 1000 試合所要時間

### 5.3 snowl と多回数ランダムサンプリングプレイヤの比較

多回数ランダムサンプリングプレイヤと標準の snowl との比較を行った．ランダムサンプリングプレイヤは，snowl 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．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．

プレイアウト回数に対する対 snowl のランダムサンプリングプレイヤの得点率の変化を図 1 に示し，ランダムサンプリングプレイヤの計算時間を図 2 に示す．計算時間はプレイアウト回数に対してほぼ線形に増えている．

得点率を見ると snowl の規定回数と近い 2,000 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が snowl より強くなると言える．

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と言えるからである．

よってどのサンプルを取って評価しても，ある最良手 $j$ の評価積算にほとんど重大な影響を及ぼしていないことから，コンピュータ大貧民が単貧民同様に，低い偶然手番感度をもっているといえる．

ところで，消費時間を同じとすると，ランダムサンプリングプレイヤは 7000 回のシミュレーションが可能であり，得点率は snowl より高い．これは，現状のままでも，推定をしないモデルの方が実戦において強かった可能性をしめしている．

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，コンピュータ大貧民プレイヤを用いた実験では，53 枚の実ゲームにおいても，やはり偶然手番感度が高くないことを示唆する結果を得た．

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け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同じ速さで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ときの，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唆された．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．

### 参考文献

[1] 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).

[2] 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).

[3] 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).

[4] 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).

[5] Luckhardt, C. A. and Irani, K. B.: An algorithmic solution of N-person games, *AAAI-86*, pp. 158–162 (1986).

[6] Sturtevant, N. and Korf, R.: On Pruning Techniques for Multi-Player Games, *Proceedings AAAI-2000* (2000).

[7] 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「snowl」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).

[8] 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).

[9] 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).

[10] Auer, P., Bianci, N. C. and Fischer, P.: Finite-time Analysis of the Multiarmed Bandit Problem, *Machine Learning*, Vol. 47, pp. 235–256 (2002).

[11] Humter, D. R.: MM algorithms for generalized Bradley-Terry models, *The Annals of Statistics*, Vol. 32, No. 1, pp. 384–406 (2004).

[12] Coulom, R.: Computing Elo Ratings of Move Patterns in the Game of Go, *ICGA Journal*, Vol. 30, No. 4, pp. 198–208 (2007).