3 不完全情報ゲーム

西野哲朗（電気通信大学大学院情報理工学研究科）

不完全情報多人数ゲーム研究の現状

不完全情報ゲームとは，麻雀やカードゲームなどのように，プレイヤごとに得られるゲームの状態に関する情報が部分的で不完全なゲームである．ゲームの状態が分からないとモンテカルロ木探索などの探索的手法は適用できない．そこで，可能な状態集合からランダムに状態をサンプリングし，完全情報ゲームと同様にモンテカルロ木探索を実行することが行われている．このような手法をモンテカルロサンプリングという．

一方，麻雀や大貧民など，早い者勝ちで展開型の多人数ゲームの場合には，ゲーム木における利得を一対比較できないため，合理的な方法では着手を1つに決めることのできないケースが多数存在し，2人ゲームで有効であった min-max 探索が安定的に行えないことが知られている．

近年，囲碁などの大きな探索空間を持つゲームにおいてモンテカルロ木探索が高い効果をあげている[3]．完全情報2人ゲームの1つである囲碁では，Crazy Stone の登場によりプレイヤプログラムの強さが飛躍的に向上した．Crazy Stone は探索木にモンテカルロ法を用いたプレイヤプログラムである．Crazy Stone と同じように，探索木にモンテカルロ法を用いる手法を採用しているプレイヤプログラム MoGo は，囲碁のプロ棋士に勝利した．

その一方で，不完全情報ゲームにおいてもモンテカルロ木探索が有効であることが分かってきているが，その理論的説明はなされていない．Long らは，不完全情報ゲームを大きく2つに分類している[6]．1つはトリック型ゲームと呼ばれ，ターンごとにカードを見せ合うことで，ゲームが進むにしたがって徐々に情報が明らかになるゲームである．大貧民もこのタイプである．もう1つはポーカー型ゲームで，勝敗決定までのプロセス内では明らかな情報開示のないゲームである．麻雀も情報開示が進む意味では大貧民と同じくトリック型ゲームに近いが，非開示の手札が 13 あり，見えない山札も平均的に 50 以上と多いためポーカーに近い面も持っている．このような両者の中間的なゲームともいえる麻雀においては，モンテカルロ木探索によって良好な結果が得られている．

大貧民を含むトリック型ゲームでは，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．しかしながら実験的には必ずしも有効と言えないゲームもあることが報告されている[4]．これまで，モンテカルロ法をゲームのプレイヤプログラムに応用する方法が模索されてきた結果，プレイヤプログラムが人間に完全情報2人ゲームでは勝つことができるようになったが，不完全情報多人数ゲームでは，いまだ人間に勝つようなプレイヤプログラムは開発されていない．

不完全情報多人数ゲームの1つである大貧民で，須藤らは，モンテカルロ法と，その制御に UCB1-TUNED と呼ばれるアルゴリズムを用いた．そして，第4回 UEC コンピュータ大貧民大会 (UECda-2009) において優勝を収めている．須藤らが用いた UCB1-TUNED とは多腕バンディット問題を解決するためのアルゴリズムである．

この UCB1-TUNED とは異なる考えに基づいた多腕バンディット問題を解決するためのアルゴリズムに，ϵ-GREEDY と呼ばれるアルゴリズムが

ある[1]．そこで本稿では，コンピュータ大貧民に対して，モンテカルロ法の制御に ϵ-GREEDY を用いた強力なプレイヤプログラムを紹介する．

コンピュータ大貧民

UEC コンピュータ大貧民大会（UECda）を，毎年 11 月末に電気通信大学で開催している[5]．本大会ではプログラム同士の高速対戦を行うため，配布されたカードの善し悪しに左右されない，プレイのアルゴリズム本来の優劣を競うことができる．

大貧民はトランプで遊ぶカードゲームの 1 つで，「ど貧民」，「大富豪」，「階級闘争」などとも呼ばれる．カードを参加者にすべて配り，手持ちのカードを順番に場に出して早く手札をなくすことを競うゲームである．1 ゲームでの順位が次ゲーム開始時の有利不利に影響する点が特徴で，勝者をより有利にするゲーム性から大富豪との名称がついた．

地方ルールが数多く存在することも大きな特徴である．地方ルールには，一度負け出すとなかなか逆転できないという欠点を補正する方向に働くものが多い．順位は，手持ちのカードのなくなった順に，大富豪，富豪，平民，貧民，大貧民となる（平民は複数存在し得るが，存在しない場合もある）．第 2 ゲーム以降は，カードを配った後のゲーム開始時までに，大貧民は大富豪に 2 枚，貧民は富豪に 1 枚，手持ちの最も強いカードを差し出さなければならない．このカード交換を「税金」または「献上」という．

トランプの大貧民は，日本発祥のゲームである．ルールがシンプルで多くの日本人が知っているゲームだが，その割に，奥が深く，地方ルールなどもたくさんある．おそらく必勝手がなく，名人やグランド・マスターもいないという特殊なゲームである．

上記大会で採用している大貧民のルールは，以下の通りである．

- **ゲームの開始**：ゲームはダイヤの 3 を持っている人から始まる．ただし，必ずしもダイヤの 3 を出さなくてもよい．
- **パスについて**：場のカードと手札の関係上，カードを出せない場合はパスとなる．カードが出せる場合でも戦略上パスすることができるが，いったんパスすると，場が流れるまで自分に順番が回ってくることはない．
- **スペードの 3**：スペードの 3 はジョーカーよりも強い．ジョーカーが 1 枚で出された場合，スペードの 3 で切ることができる．
- **場の流れ方**：全員がパスしたら場が流れ，最後にカードを出した人が場にカードがない状態からカードを出すことができる．仮に自分以外がパスしたとき，自分がカードを出すことができれば連続してカードを出すことができる．
- **8 切り**：8 を含んだ手を出した場合，場のカードがクリアされカードを出した人が任意のカードを出すことができる（権利をとることができる）．
- **革命**：同じ番号のカードを 4 枚，もしくはジョーカーを含んだ 5 枚をセットで出すと，革命が起こる．革命後はカードの強さが逆転する．
- **階段（シークエンス）**：同一マークの連番が 3 枚以上ある場合は，同時に出すことができる．5 枚以上同時に出すと革命が起こる．
- **しばり（ロック）**：場にあるカードと同じマークのカードを出すと「しばり」状態となり，以後同じマークしか出せない．
- **あがり方**：どんなカードでもあがることができる．
- **カードの交換**：大貧民は大富豪に 2 枚，貧民は富豪に 1 枚，それぞれ強いカードを献上する．逆に，大富豪は 2 枚，富豪は 1 枚，カードを返す．その選び方は任意で，強いカードを返してもよい．なお，上記大会のシステムでは，これらのカードは自動的に選択される．

本大会で使用したプログラムは，カードの配布や場の管理を行うサーバ・プログラムと，プレイヤに対応するクライアント・プログラムから構成される（図 -1 参照）．5 人のプレイヤに対応する 5 つのクライアント・プログラムを，サーバ・プログラムにつないで対戦を行う（図 -2，3 参照）．非常にシンプルな大貧民クライアントの構成例を，図 -1 ～図 -4 に流れ図の形で示す．サーバ，標準クライア

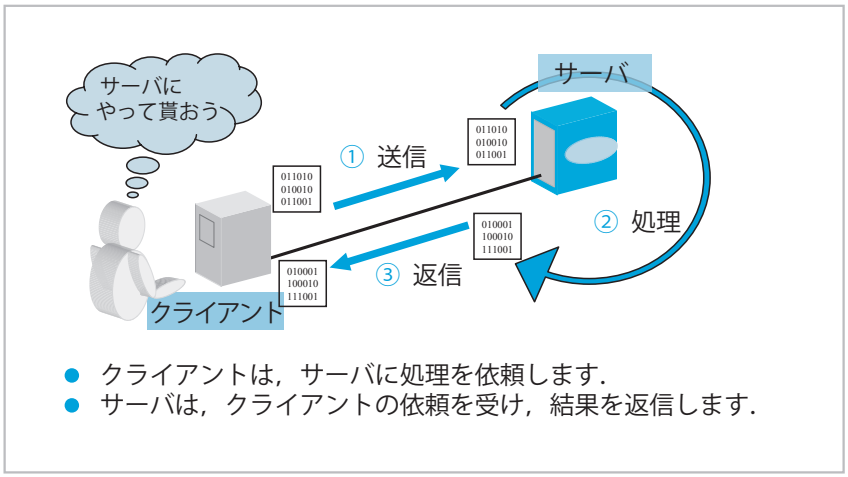

図 -1
サーバ—クライアント・システム

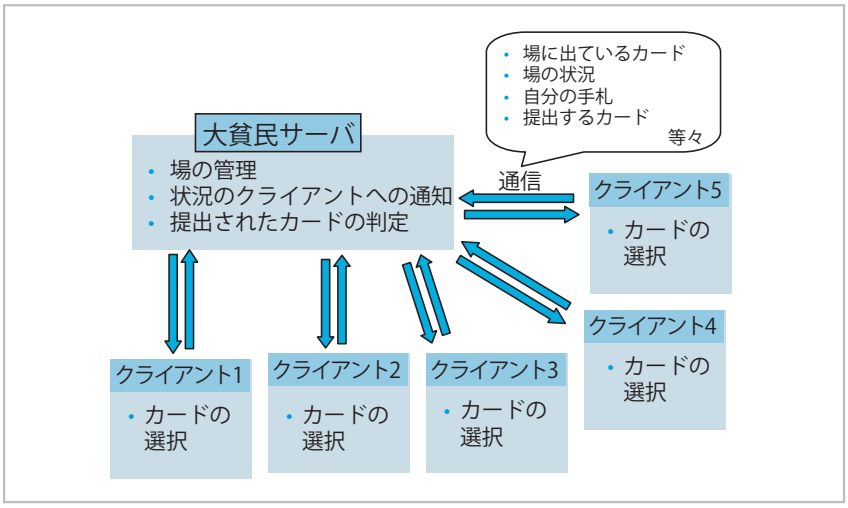

図 -2
システム構成図

ント等のプログラムのソース・コードは，大会サイトからダウンロード可能である．

大貧民に対するアルゴリズム

● ϵ-GREEDY

多腕バンディット問題を解決するためのアルゴリズムの 1 つとして，ϵ-GREEDY が知られている [1]．筆者らは，コンピュータ大貧民において，ϵ-GREEDY をモンテカルロ法におけるプレイアウトの制御アルゴリズムとして用いることを提案した [2]．ϵ-GREEDY による n 回目のプレイアウトを行う合法手の選択は，以下のように行う．

1. 確率 $1-\epsilon_n$ で，$n-1$ 回目までの平均報酬値 $\overline{X}_i$ が最も大きい合法手を選択する．
2. 確率 ϵ_n で全 K 個の合法手の中からランダムに選択する．

ここで，ϵ_n は式 (1) で計算される．

$$\epsilon n = \min\left\{1, \frac{cK}{d^2 n}\right\} \qquad (0 \leq \epsilon n \leq 1) \qquad (1)$$

この ϵ-GREEDY を用いることにより，n 回目のプレイアウトを行う合法手を選択する際，$\overline{X}_i$ が最も大きい合法手を選択する確率は $1-\epsilon+\frac{\epsilon}{K}$ となる．そして，他の合法手を選択する確率はそれぞれ $\frac{\epsilon}{K}$ となる．このようにして，各局面において最善手である可能性が高い合法手に，多くのプレイアウトを行うことができる．一方，一定の確率を他の合法手にも割り当てることで，各合法手に対しても探査を行うことができる．

上記のパラメータ c は，$c=\frac{N}{K}$ と取るとプレイ

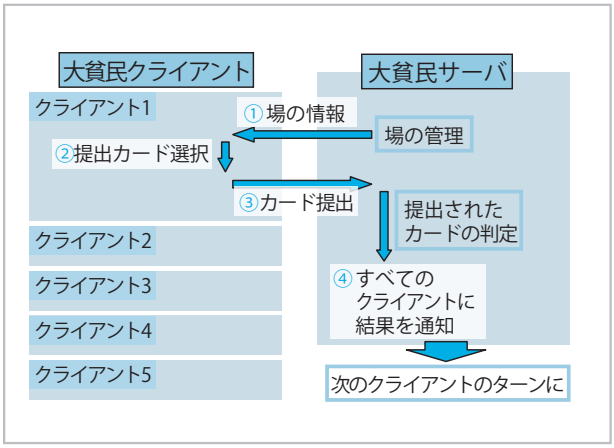

図 -3　処理の流れ

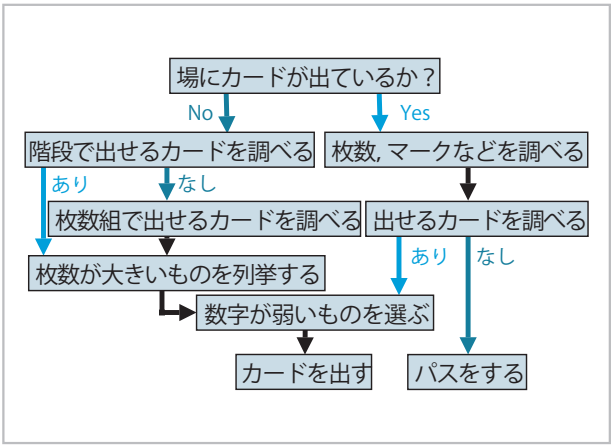

図 -4　提出カード選択のフローチャート

アウトを行う合法手をランダムにしか選択できなくなる．また，$c=0$ と取ると，$\overline{X_i}$ が最大の合法手しか選択できなくなる（貪欲法）．このことから，パラメータ c を調節することで，対象となる問題に適した情報を得ることができると考えられる．さらに，パラメータ d を変更することでも種々の調整が可能だが，ϵ-GREEDY のパラメータによる性能差は，d よりも c に因るところが大きい．そのため，暫定的には，$d=1$ としておけば十分である．

モンテカルロ法による合法手の探査に ϵ-GREEDY を用いたコンピュータ大貧民のプレイヤプログラムを，以下では epsilon と呼ぶ．

● 成功確率の増幅

epsilon により，ゲームにおける各局面の最善手を導き出す際には，プレイアウトを行う回数を増やし，ゲームのルールに反しない限り，シミュレーションに時間を費やすことが理想的である．しかし，乱数を用いた制御アルゴリズムでは，同じ局面に対して同じ合法手を最善手として推定するとは限らない．すなわち，ある局面に対して K 個の合法手の中に最善手がただ 1 つ存在すると仮定すると，局面における最善手を推定する確率（成功確率）が $\frac{1}{K}$ である可能性も考えられる．

そこで，epsilon による最善手の推定を 5 回行わせ，そのうち最善手と推定された回数の最も多い合法手を最善手とすることで，成功確率を増幅させる手法を採用した．具体的には，α という合法手を最善手として 3 回推定し，β という合法手を最善手として 2 回推定した場合，推定結果を α とする．この手法は乱数を用いたアルゴリズムにおける成功確率を増幅する基本的な考えであり，成功確率が $\frac{1}{2}$ よりも大きければ成功確率を飛躍的に高められることが知られている．

このように実装したコンピュータ大貧民のプレイヤプログラムを以下では majority と呼ぶ．このように最善手を推定しても，最終的な最善手を求めた際のプレイアウト回数は epsilon，majority ともに N であるため，強さを直接比較することができる．

● ランダムサンプリング

epsilon は，ϵ_n を変化させてプレイアウトを行う合法手を選択するプレイヤプログラムである．しかし，epsilon が，局面における最善手を推定する上で有用であるかは分からない．そこで，全 K 個の合法手の中から合法手をランダムに選択し，プレイアウトを行う合法手を選択する場合（ランダムサンプリング）と epsilon との強さの比較を行った．これにより，ϵ-GREEDY の有用性を検証できると考えられる．なお，このように実装したコンピュータ大貧民のプレイヤプログラムを以下では random-sampling と呼ぶ．

● soft max 関数の応用

epsilon では，$\overline{X_i}$ が最大の合法手のみに，他の合法手よりも大きな確率 $1-\epsilon+\frac{\epsilon}{K}$ が与えられる．そして，他の合法手に関しては，$\overline{X_i}$ がどの程度小さいかは考慮されず，確率 $\frac{\epsilon}{K}$ が与えられる．そこで，$\overline{X_i}$ の大きさに応じて，合法手 i を選択する確率 P(i) を与えるために式 (2) で表される soft max と呼ばれる関数を導入し比較を行った．soft max 関数を用いて実装したコンピュータ大貧民のプレイヤプログラムを以下では soft-max と呼ぶ．

$$P(i)=\frac{e^{\frac{\overline{X_i}}{r}}}{\sum_{j=1}^{K} e^{\frac{\overline{X_j}}{r}}} \qquad (2)$$

soft max 関数で用いられている r は正定数である．この soft max 関数は，$r\to 0$ の極限では貪欲法と動作が一致し，$r\to\infty$ の極限ではランダムサンプリングと動作が一致することが知られている．

UEC コンピュータ大貧民大会の公式ルールでは，1 位から順に 1 試合につき 5，4，3，2，1 点が与えられる（プレイヤ数は 5）．以下では，このルールに基づき，$\overline{X_i}$ を 1 以上，5 以下の実数とした．このため，パラメータ r を 0 より大きく，かつ，5 以下の実数とした．以上より，$r\to 0$ で貪欲法と同じ動作をし，$r=5$ でランダムサンプリングに近い動作をすることとなった．

● 対戦による強さの比較実験

ϵ-GREEDY と soft max 関数のどちらが優れているかは，対象とする問題と密接にかかわっているとされ，いまだ詳しくは知られていない．したがって，epsilon と soft-max の強さを比較することで，ϵ-GREEDY と soft max 関数のどちらが，コンピュータ大貧民に対してモンテカルロ法を用いる際に，有効であるのかを調べることとした．

epsilon のパラメータ c と soft-max のパラメータ r を変化させ，パラメータの変化が強さとどのような関係があるかを調べた．なお，5 つのプログラムでゲームを行わなければならないため，

- 4 つの UCB1-T との対戦
- 4 つの random-sampling との対戦
- 4 つの default（UECda 標準プレイヤプログラム）との対戦

を行った．ここで，UCB1-T は先行研究により実装された，モンテカルロ法に UCB1-TUNED を応用したプレイヤプログラムである．

これらのプレイヤプログラムからどれだけの点数を獲得できるかで epsilon と soft-max における最も適したパラメータを求める予備実験を行った．なお，プレイアウトを行う総回数 N は，先行研究において 1500 回とされていた．そこで，すべてのプレイヤプログラムにおいて N=1500 として実験を行うこととした．また，実験では UECda で用いられているルールに基づき実験を行った．

予備実験の後，最適なパラメータを用いた epsilon，soft-max，majority と UCB1-T，random-sampling の 5 つのプレイヤプログラムを同時に対戦させて，最終的な強さの比較を行った．

予備実験の結果，epsilon のパラメータ c の最適値は $c=\frac{1500}{2K}$ となり，soft-max のパラメータ r の最適値は r=1 となった．これらのパラメータを用いた epsilon，soft-max，majority と UCB1-T，random-sampling を対戦させたところ，1 位から順に majority，epsilon，UCB1-T，random-sampling，soft-max となり，majority が他のプレイヤプログラムより群を抜いて強いことが示された [2]．

将来展望

最近，不完全情報ゲームに対するモンテカルロ法の適用について，さまざまな研究が行われるようになってきた．そのような研究の具体的な事例として，本稿では，モンテカルロ法におけるプレイアウトの制御に ϵ-GREEDY を用いた，コンピュータ大貧民のプレイヤプログラム epsilon を紹介した．本プレイヤプログラムは，モンテカルロ法のプレイアウトを制御する random-sampling や soft-max を用いたプレイヤプログラムよりも多くの点数を獲得でき

た．以上のことより，コンピュータ大貧民では，モンテカルロ法のプレイアウトの制御に ϵ-GREEDY を用いることが有効であることが示された．

さらに，プレイヤプログラム epsilon が最善手を推定する確率（成功確率）を高めるために，推定回数を 5 回に増やしたプレイヤプログラム majority を紹介した．majority は epsilon に勝利することができたが，これは，ϵ-GREEDY は推定回数を増やすことで成功確率を高めることができることを示している．

このような研究を足がかりとして，他の不完全情報ゲームに対するより強力なプレイヤプログラムを開発していくことが今後の課題となる．

参考文献

1) Auer, P., Cesa-Bianchi, N. and Fischer, P.：Finite-time Analysis of the Multiarmed Bandit Problem, Machine Learning, 47: pp.235-256 (2002).
2) 小沼　啓，西野哲朗：コンピュータ大貧民に対するモンテカルロ法の適用，情報処理学会第 25 回ゲーム情報学研究会資料集 (2011).
3) 美添一樹：モンテカルロ木探索―コンピュータ囲碁に革命を起こした新手法―，情報処理，Vol.49, No.6, pp.686-693 (June 2008).
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) 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，Vol.48, No.8, pp.884-888 (Aug. 2007).
6) Long, J., Sturtevant, N. R., Buro, M. and Furtak, T.：Understanding the Success of Perfect Information Monte Carlo Sampling in Game TreeSearch, Proceedings of the 24th. AAAI Conf., AAAI, pp.134-140 (2010).

(2011 年 11 月 14 日受付)

西野哲朗（正会員）
nishino@ice.uec.ac.jp

昭和 57 年早稲田大学理工学部数学科卒業．昭和 59 年同大学院理工学研究科博士前期課程修了．日本アイ・ビー・エム (株)，東京電機大学，北陸先端科学技術大学院大学を経て，平成 6 年電気通信大学電気通信学部電子情報学科助教授．平成 18 年同大同学部情報通信工学科教授．平成 22 年同大学院情報理工学研究科総合情報学専攻教授．現在に至る．理学博士．平成 7 年本会 Best Author 賞，平成 10 年人工知能学会研究奨励賞，平成 14 年電子情報通信学会ソサイエティ論文賞，平成 15 年船井情報科学振興賞，平成 20 年 IBM Faculty Award, 平成 22 年文部科学大臣表彰科学技術賞，平成 23 年モノづくり連携大賞特別賞各受賞．