

不完全情報ゲームにおける MCTS へのグループ化の適用

坪倉 弘治^{1,a)} 西野 順二^{1,b)}

概要: 囲碁や将棋などのゲーム AI で優れた成績を示したモンテカルロ木探索は節点の分岐数が多い場合に性能が落ちてしまう。そのようなゲームとして不完全情報ゲームなどがある。不完全情報ゲームでは、確定できない情報についてすべての可能性を考慮しなければならないため、節点の分岐数が増え探索空間が非常に大きくなる。この問題を解決する手法としては Information Set MCTS などが存在する。本研究では分岐数の減少のため、節点のグループ化を利用する。本研究では情報が不完全であるがゆえに考慮する手の数が増える不完全情報ゲームにおいて、不完全情報ゲームの大貧民を実験対象ゲームとしてグループ化の効果を検証し、異なるグループ化の方法をとった AI の比較・評価を行った。実験により大貧民で効果があることを示した。

Grouping for MCTS in Imperfect Information Games

TSUBOKURA KOJI^{1,a)} NISHINO JUNJI^{1,b)}

Abstract: Monte Carlo tree search, which has shown excellent performance in Go and Shogi, performs poorly when the number of branches of nodes is large. Such games include incomplete information games. In incomplete information games, the number of branches of nodes increases and the search space becomes very large because all possibilities must be considered for information that cannot be determined. Information Set MCTS is one of the methods to solve this problem. In this study, we use Grouping Nodes to reduce the number of branches. And, we tested the effect of Grouping in imperfect information game, Daihinmin, in which the number of moves to be considered increases due to incomplete information, and compared and evaluated AI with different grouping methods. Experiments show that it works in Daihinmin.

1. はじめに

囲碁や将棋などのゲーム AI で優れた成績を示した AI アルゴリズムの一つがモンテカルロ木探索 (Monte Carlo Tree Search; MCTS)[1] である。この手法は二人確定完全情報ゲームだけでなく、多人数ゲームや不完全情報ゲーム、不確定情報ゲームにおいても用いられている [2] が、探索は困難である。探索が困難である理由の一つが、節点の分岐数が多い場合にモンテカルロ木探索の性能が落ちてしまうという問題である。そして、根節点の分岐数が非常に多い場合には、一手目の節点にすら探索に掛からない、一度も訪問されない節点が見れる。その場合、訪問されない節

点は最適手を決定する際に最適かどうか分からないため、最適であったとしても選択されなくなってしまう。そのような節点の分岐数が多く探索空間が広大なゲームとして、不完全情報ゲームや不確定情報ゲーム・ターン制戦略ゲームなどがある。

不完全情報ゲームにおいては、確定できない情報についてすべての可能性を考慮しなければならないため、探索空間が非常に大きくなる問題がある。そのため、広い探索空間を効率的に探索するアルゴリズムが必要となる。この問題を解決する手法としては Information Set MCTS[3] などが存在する。

本研究では分岐数の減少のため節点のグループ化を利用する。グループ化はターン制戦略ゲーム TUBSTAP の AI、M-UCT[4] や S-UCT[5] でその有効性が示されており、分岐数が多く探索空間が広いゲームで探索の効率化を行える

¹ 電気通信大学大学院
The University of Electro-Communications
^{a)} tsubokura.koji@mail.uec.jp
^{b)} nishinjunji@uec.ac.jp

手法である [6]。

本研究では情報が不完全であるがゆえに考慮する手の数が増える不完全情報ゲームにおいて、不完全情報ゲームの大貧民を実験対象ゲームとしてグループ化の効果を検証し、異なるグループ化の方法をとった AI の比較・評価を行った。

2. 不完全情報ゲームの MCTS

2.1 モンテカルロ木探索

モンテカルロ木探索 (Monte Carlo Tree Search; MCTS) は、モンテカルロ法に木探索の考えを取り入れた探索法であり、2006 年に R. Coulom によって提唱され、[1][7] 同氏が作成した囲碁プログラム「Crazy Stone[1]」や、「MoGo[8]」などで使用されている。モンテカルロ木探索は、囲碁などの評価関数の作成が困難だが、簡単にゲームの勝敗判定ができるというゲームにおいて有効なアルゴリズムであり、探索空間が大きい場合でもおおむね効果的であることから、不完全情報ゲームや不確定情報ゲームにおいても用いられている。

モンテカルロ木探索では、プレイアウトを繰り返し行い、節点の価値を近似的に求める。プレイアウトには次の 4 段階がある。

- (1) 木探索: 何らかの方法で根節点 v_0 から展開していない節点 (先端節点) v_l まで探索木を下る。
- (2) 展開: v_l のプレイアウト回数が既定値を超えている場合に v_l を展開し、 v_l の各子節点についてランダムシミュレーションとバックアップを行う。
- (3) ランダムシミュレーション: v_l の状態 s_l からゲームの終端までランダムに着手し、 v_0 のターンプレイヤー p から見た利得 Δ を得る。
- (4) バックアップ: ランダムシミュレーションで得られた結果を元に v_l とその先祖節点の値を更新する。

ここで v_l のプレイアウト回数とは、木探索における v_l の訪問回数のことである。

ゲームの利得については、ゲームの勝ちを 1、引き分けを $\frac{1}{2}$ 、負けを 0 とすることが多い。

2.1.1 UCT 探索

多腕バンディット (Multi-Armed Bandit) 問題の解法の一つである UCB1 アルゴリズム [9] を、モンテカルロ木探索の木探索で下る節点の選択方法 (Tree Policy) に使用したのが UCT (UCB applied to Trees) 探索 [10] である。この探索法は Kocsis, Levente and Szepesvári, Csaba らによって 2006 年に提唱された。

このアルゴリズムでは UCB 値が最も高い子節点が下る節点として選択される。節点 j の子節点からバックアップされた利得の平均 (平均利得) を \bar{x}_j 、プレイアウト回数を n_j 、親節点のプレイアウト回数を n とし、UCB 値は以下のように定義される。

$$ucb = \bar{x}_j + C \sqrt{\frac{2 \log n}{n_j}} \quad (1)$$

定数 C は問題に合わせて調整される。 x の範囲が $[0, 1]$ であるとき、理論上 $C = 1$ が良い。UCT 探索は十分な探索時間が与えられたとき、探索木がネガマックス木に収束する。[11]

2.2 ISMCTS

不完全情報ゲームに MCTS を適用するに当たり、探索空間が増大する問題に対処したものが Information Set MCTS (ISMCTS)[3] である。ISMCTS では探索木の節点をゲームの状態ではなく情報集合に対応付け、探索空間を小さくすることでこの問題の解決を図っている。提案論文では ISMCTS の 3 つの形式が示されており、それは次のようなものである。

- (1) Single-Observer ISMCTS (SO-ISMCTS): ISMCTS の基本となる形式。
- (2) SO-ISMCTS+POM: 決定化を用いずに相手の部分的に観測可能な手 (Partially Observable Moves; POM) をランダムな手を仮定する。
- (3) Multiple-Observer-ISMCTS (MO-ISMCTS): 異なる情報を持ったプレイヤー毎に探索木を作成する。

部分的に観測可能な手 (Partially Observable Moves; POM) とは、あるプレイヤーのとした手の情報が他プレイヤーに対し一部伝えられない手のことである。SO-ISMCTS は部分的に観測可能な手を含んだゲームで上手く働かない場合があるため、そのような場合に対応するため考案されたのが SO-ISMCTS+POM と MO-ISMCTS である。

今回実験対象とするゲームである大貧民に部分的に観測可能な手は存在しないため、本研究では元となるモンテカルロ木探索 AI として SO-ISMCTS を選択し、SO-ISMCTS に加えて節点のグループ化を行うことでさらなる分岐数の減少を目指す。

3. 節点のグループ化

3.1 不完全情報ゲームの節点

不完全情報ゲームとは、カードゲームの手札・山札やジャンケンで相手の出す手など、プレイヤー毎に得られる状態や行動の情報が限られたゲームのことである。情報が不完全であることにより、プレイヤーが現在までのプレイで得られた情報から現在のゲームの状態を区別できなくなることがある。この互いに区別できなくなる状態の集合を情報集合といい、プレイヤーは情報集合のある状態に到達したことを知ることができるが、情報集合のどの状態に実際に到達したかは知らない。

木探索を行うモンテカルロ木探索では、探索木を形成するために合法手や次の状態を用意しなければならないが、

情報集合の存在によってプレイヤーは情報集合のどの状態に到達したかを知ることができないので、情報集合のすべての状態について合法手や次の状態を考慮しなければならない。したがって、完全情報ゲームでは1つの状態について考えればよかったことが、不完全情報ゲームでは考える対象が情報集合の全状態に増えるため、探索木の分岐数が増加し探索空間が広がる。

3.2 節点のグループ化

節点のグループ化とは、探索木の中の注目する節点の子節点をいくつかのグループに分け、各グループ毎にグループ節点を作成して親節点と子節点の間に挿入することである [6]。これにより、下る節点を選択する前にグループにまとめた節点を選択するため探索木の分岐因子が減少する。グループ化はターン制戦略ゲームのような分岐数が多く探索空間が広いゲームで探索の効率化を行うことができ、それは探索で行えるプレイアウトの回数に対し相対的に探索木の分岐因子が多い場合に顕著である [6]。同じグループにまとめる節点は節点の価値が近いものであればグループ化の効果が現れることが分かっている。

4. 大貧民におけるグループ化の検証

4.1 大貧民のルール

多人数不完全情報ゲームである大貧民を対象とし検証実験を行う。大貧民には多数のローカルルールが存在するが、本研究では UEC コンピュータ大貧民大会の標準ルール [12] を用いた。そのうち本研究に大きな関係があるルールを抜粋して以下に示す。

カードの出し方

場にカードがある場合、同じ枚数・出し方(ペア・階段・縛り)でより強いカードしか場に出せない。場にカードがない場合、任意の出し方ができる。場にカードがない場合に任意のカードを場に出せるため、場にカードが有る場合より分岐数が多くなる。

ジョーカー

使用するカードはトランプ 52 枚とジョーカー 1 枚で、ジョーカーは任意のカードの代わり、またはペア出しの 5 枚目・階段出しの 3 より弱い・2 より強いカードとして使用できる。ジョーカーが手札にあるかないかで分岐数が大きく変化する。

8 切り

場に出すカードに 8 が含まれるなら場を流しもう一度好きなカードを場に出せる。これにより次の自分の手番で場が空のため、次の節点の分岐数が多くなる。

4.2 大貧民における分岐数

大貧民の平均分岐数はおよそ 6 であると知られている。大貧民においてはパスしかできない、もしくは 2 つ 3 つの

行動しかできないような手番は多いが、数は少ないものの分岐数が非常に多い節点も存在する。分岐数が極端に少ない節点では探索を行う必要性は低く、可能な手が 1 つしかないような節点は探索する必要がない。それに対して分岐数が非常に多い節点は探索を行う必要性が高く、そのような節点で十分な探索ができるかが AI の性能に寄与すると考えられる。

大貧民において特に分岐数が多く最適行動の選択が難しい状態は空の場で手番が回ってきた状態である。この状態では任意のカードの出し方ができるため、他の状態より分岐が増える。手札の情報がわからない他プレイヤーが手番プレイヤーだった場合には、自分の手札に所持していないカードから出し方を考慮する必要があるためさらに分岐が増える。手番プレイヤーは他プレイヤー、カードの全体がトランプ 52 枚とジョーカー 1 枚の 53 枚、手札の枚数が 11 枚で全て未知として、ジョーカーを使わない場合の分岐数 n は

$$n \leq 435$$

となる。ジョーカーが使える場合にはジョーカーが任意のカードの代わりとして使用できるため分岐数が増え、分岐数 \tilde{n} は

$$\tilde{n} \leq 2205$$

となり、場に何らかのカードが出ている場合の分岐数と比べると非常に多い。

グループ化の効果はモンテカルロ木探索で行えるプレイアウトの回数に対し相対的に探索木の分岐因子が多い場合に顕著であるため、空の場の節点で行動についてグループ化を行うことにより分岐因子を減らし、探索の効率化ができると考えられる。

4.3 グループ化の方法

グループ化は場が空のときの節点のみ行うこととし、それ以外の節点ではグループ化を行わず通常の ISMCTS と同様に節点を展開する。グループ分けの基準は次のような 3 つの基準を用意し、効果を比較する。

- (1) 手札の出し方(枚数とペア/階段)についてグループを分ける。ただしパス行動については 1 枚出し(以下 1 枚ペアと呼ぶ)と同じグループに入れる。
- (2) (1) と同様ではあるが、パス行動を 1 枚ペアとは別に分ける。
- (3) 出し方の枚数についてグループを分ける。パス行動は 1 枚出しと同じグループに入れる。

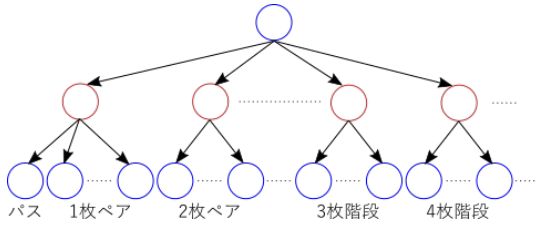


図 1 (1) の模式図

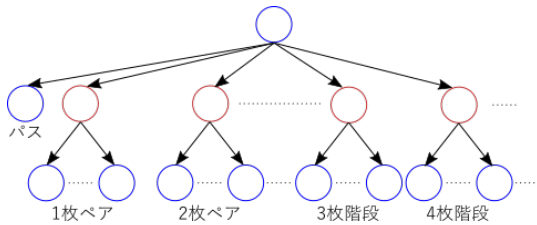


図 2 (2) の模式図

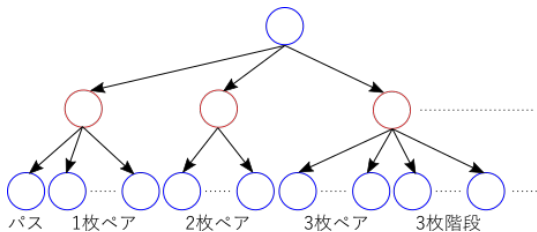


図 3 (3) の模式図

4.4 グループ化の効果比較実験

グループ化によりモンテカルロ木探索の性能が向上するかを検証する実験として、ISMCTSのAI(ISMCTS)とISMCTSに節点のグループ化を追加したAI(AI1,2,3)の対戦を行った。実験の対象ゲームは大貧民とし、実験でのプラットフォームにUECda[13]で公開されているものを使用した。また、大貧民のルールはUECdaで定められたUEC標準ルール[12]を使用した。

UECdaの大貧民は参加プレイヤー数が5人であるため、比較を行うAI2つに加え、UECdaで用意されているルールベースのデフォルトAI(Default)を3つ参加させたゲームを1000回実行するのを、デフォルトAI3つを固定し比較するAI2つを入れ替え、総当りで繰り返し行った。一戦終了時に先に上がったプレイヤーから順に1, 0.75, 0.5, 0.25, 0の[0, 1]範囲のスコアが与えられ、各1000ゲームの平均スコアを求めた。

ISMCTSのTree PolicyはUCTとし、モンテカルロ木探索・UCTの各種パラメータは次のように設定した。

終了探索回数	展開しきい値	UCBのC値
1000	10	1.0

4.5 結果

実験結果を以下の表2に示す。ただし、表の各項に示される数値は対応する行・列のAIの対戦結果で、同じ行ラベルのAIが獲得したスコアの合計が示されている。表の上三角と下三角の同じAIの組み合わせに対応する項は同じ対戦の結果で、もう片方のAIが獲得したスコアの合計が示されている。対戦に参加したデフォルトAIの獲得したスコアに関しては省略している。

表 2 対戦結果-合計スコア

	AI1	AI2	AI3	ISMCTS
AI1	-	612.75	635.75	571.50
AI2	539.25	-	523.75	571.75
AI3	595.25	602.50	-	604.00
ISMCTS	537.75	565.00	543.75	-

各対戦における大貧民から大富豪までの各順位の獲得回数を以下の表3から8までに示す。

表 3 各順位獲得数-AI1vsAI2

	大貧民	貧民	平民	富豪	大富豪
AI1	124	120	204	285	267
AI2	152	214	196	201	237

表 4 各順位獲得数-AI1vsAI3

	大貧民	貧民	平民	富豪	大富豪
AI1	125	167	209	262	293
AI3	153	189	202	260	252

表 5 各順位獲得数-AI2vsAI3

	大貧民	貧民	平民	富豪	大富豪
AI2	154	206	213	245	182
AI3	140	135	197	231	297

表 6 各順位獲得数-AI1vsIS

	大貧民	貧民	平民	富豪	大富豪
AI1	137	169	211	237	246
ISMCTS	188	162	197	217	236

表 7 各順位獲得数-AI2vsIS

	大貧民	貧民	平民	富豪	大富豪
AI2	129	193	197	224	257
ISMCTS	165	145	205	235	250

表 8 各順位獲得数-AI3vsIS

	大貧民	貧民	平民	富豪	大富豪
AI3	117	152	200	260	271
ISMCTS	164	170	218	223	225

表2の結果を見ると、AI1がAI2の合計スコアを上回り、AI1がAI3の合計スコアを上回り、AI3がAI2の合計スコアを上回り、AI1から3すべてがISMCTSの合計スコアを上回っていることがわかる。表3から8までの各順位の獲得回数について自由度4のカイ二乗検定を行った結果、すべての対戦結果について有意水準 $\alpha = 0.005$ で有意差が存在する。

5. 大貧民におけるグループ化の効果

AI1からAI3すべてがISMCTS、つまり元のISMCTSよりも有意に上回っていることから、不完全情報ゲームにおいてもグループ化で分岐数を減少させることで探索の効率化を行うことができると考えられる。

また、グループ化の方法によって性能に差が現れることもわかった。まず、AI1がAI2を合計スコアで上回っていることから、パスを1枚ペアのグループに入れる(AI1)か入れない(AI2)かではグループに入れるほうがより良いと言える。この理由としては、グループ化を行っている場が

空の状態であえてパスをしなければならないような状況が非常にまれであることが考えられる。他のプレイヤーより先に手札の枚数を0にするのが目的となる大貧民において、相手の行動を制限しつつ自分の手札を減らせる行動をとらず、そのような行動ができる場が空という状態を他プレイヤーに渡してしまうパスはまず取るべきでない行動である。当然 AI1 と AI2 の両方ともそのような最適とは見えない行動をモンテカルロ木探索の結果選択することはほとんどないが、探索時点での枝刈りは行っていないためモンテカルロ木探索のプレイアウトが割り当てられる。そして、パスを1枚ペアのグループに加える方法では1枚ペアのグループの子節点と横並びになるのに対し、パスをグループに加えない方法ではパス節点以外のグループ節点と横並びとなるため、プレイアウトの訪問回数が増えてしまう。パス行動はまず最適手ではないことを考えると、最適手が存在する他の節点を探索する回数が減ってしまうため、パスをグループに加える方法に比べて十分な探索が行えず、合計スコアが下回ったと考えられる。

AI1 は AI3 を合計スコアで有意に上回っていることから、出し方の枚数について同じグループにするのではなく、ペアと階段は別のグループに分けるほうがより良いと言える。しかし、AI1 と AI3 は他のグループ化 AI 同士の対戦に比べ、有意な差ではあるものの合計スコアの差が小さい。その理由としては1と3の方法の違いが小さいということが挙げられる。3の方法では出し方の枚数についてグループを作成しているが、これで1の方法と違いが出るのは出す枚数が3枚以上の節点のみである。UEC 標準ルールの大貧民では階段出しが3枚以上の場合に可能な出し方のため、1枚ペアや2枚ペアの節点に関しては1と3の方法でグループ化が同じである。プレイヤーの手札に3枚以上の出し方が可能な組よりも、1枚ペアと2枚ペアが可能な組の方がより多く現れやすいということを考えると、1と3の方法の違いは小さいと言える。したがって対戦結果の合計スコアの差が小さくなったと考えられる。

6. まとめ

分岐数が多くなる不完全情報ゲームにおいてグループ化の効果を検証するため、UECdaの大貧民を対象ゲームとして、ISMCTSのAIとISMCTSにグループ化を行ったAIの対戦実験を行った。グループ化を行うのは場が空の状態の節点でのみとし、カードの出し方(枚数・パス/階段)によってグループを作成した。そしてグループ化の方法の細部が異なるAIを3種類作成し、その比較・評価を行った。実験の結果、不完全情報ゲームにおいても、グループ化で分岐数を減少させることで探索の効率化を行うことができると言える。

また、グループ化の方法によりその性能に差が生じることがわかった。パスを1枚グループに加える方法はグルー

プに加えない方法より有意に性能が高く、カードの出し方の枚数が同じものを同じグループに入れる方法は入れない方法より有意に性能が低い、という結果が得られた。

参考文献

- [1] Rémi Coulom. Efficient selectivity and backup operators in monte-carlo tree search. In *International conference on computers and games*, pp. 72–83. Springer, 2006.
- [2] 三木理斗ほか. 多人数不完全情報ゲームにおける最適行動決定に関する研究. 2010.
- [3] Peter I Cowling, Edward J Powley, and Daniel Whitehouse. Information set monte carlo tree search. *IEEE Transactions on Computational Intelligence and AI in Games*, Vol. 4, No. 2, pp. 120–143, 2012.
- [4] 武藤孝輔, 西野順二. ターン制戦略ゲームにおけるuctとファジィ評価の適用. 日本知能情報ファジィ学会 ファジィシステム シンポジウム 講演論文集第31回ファジィシステムシンポジウム, pp. 226–229. 日本知能情報ファジィ学会, 2015.
- [5] 提橋凜, 西野順二. ターン制戦略ゲームにおけるユニット抽象化探索の性能. 日本知能情報ファジィ学会 ファジィシステム シンポジウム 講演論文集, Vol. 34, pp. 810–814, 2018.
- [6] 坪倉弘治, 西野順二ほか. グループ化を用いたモンテカルロ木探索の性能の分析. *ゲームプログラミングワークショップ 2019 論文集*, Vol. 2019, pp. 144–149, 2019.
- [7] Cameron B Browne, Edward Powley, Daniel Whitehouse, Simon M Lucas, Peter I Cowling, Philipp Rohlfsagen, Stephen Tavener, Diego Perez, Spyridon Samothrakis, and Simon Colton. A survey of monte carlo tree search methods. *IEEE Transactions on Computational Intelligence and AI in games*, Vol. 4, No. 1, pp. 1–43, 2012.
- [8] Sylvain Gelly, Yizao Wang, Olivier Teytaud, Modification Uct Patterns, and Projet Tao. Modification of uct with patterns in monte-carlo go. 2006.
- [9] Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, Vol. 47, No. 2-3, pp. 235–256, 2002.
- [10] Levente Kocsis and Csaba Szepesvári. Bandit based monte-carlo planning. In *European conference on machine learning*, pp. 282–293. Springer, 2006.
- [11] Levente Kocsis, Csaba Szepesvári, and Jan Willemson. Improved monte-carlo search. *Univ. Tartu, Estonia, Tech. Rep*, Vol. 1, , 2006.
- [12] UECda-2019 Project Team. Uec 標準ルールについて. http://www.tnlab.inf.uec.ac.jp/daihinmin/2019/document/_rules.html.
- [13] UECda-2019 Project Team. Uecda-2019 コンピュータ大貧民大会. <http://www.tnlab.inf.uec.ac.jp/daihinmin/2019/>.