

コンピュータ UNO におけるモンテカルロ法プレイヤーの構築

稲葉 颯弥* 大久保誠也† 若月光夫‡ 西野哲朗§
電気通信大学 静岡県立大学 電気通信大学 電気通信大学
情報理工学域 経営情報学部 大学院情報理工学研究科 大学院情報理工学研究科

1 はじめに

近年のゲーム情報学ではモンテカルロ法に関する研究が盛んに行われている。モンテカルロ法はゲーム特有の知識を用いずに対戦プログラムの実装が可能な手法である。実装の容易さから囲碁や大貧民など多くのゲームに適用され、高い有効性が実証されている [1][2]。一方、大貧民と同じ不完全情報ゲームの UNO においてはモンテカルロ法に関する研究はあまり多く行われていない。

本研究の目的は UNO におけるモンテカルロ木探索の有効性を検証することである。囲碁や大貧民で主流であるモンテカルロ木探索を実装し、性能を評価する。

2 UNO

UNO は 1971 年に Merle Robbins によって考案された、独自のカードセットを用いる不完全情報ゲームである。プレイヤーは配られた手札を減らしていき、最初に手札を 0 枚にしたプレイヤーが勝者となる。敗者は手札内容に応じた得点を引かれ、勝者は敗者が引かれた全ての得点を得る零和ゲームである。

UNO のルールは国や地域、発売年度によって大きく異なる。本研究では、NTT 東日本が主催するプログラミングコンテスト ALGORI [3] の第 1 回大会で用いられるルールを採用する。ただし、研究の都合上以下のルールは採用しない。

- 効果を自由に設定できるホワイトワイルドカード
- 残り手札が 1 枚になった時に「UNO」と宣言する

Construction of a Monte Carlo Player in Computer UNO

*Soya Inaba, School of Informatics and Engineering, The University of Electro-Communications

†Seiya Okubo, School of Management and Information, The University of Shizuoka

‡Mitsuo Wakatsuki, Graduate School of Informatics and Engineering, The University of Electro-Communications

§Tetsuro Nishino, Graduate School of Informatics and Engineering, The University of Electro-Communications

3 モンテカルロ法

3.1 単純なモンテカルロ法

モンテカルロ法は、乱数を用いた計算手法の総称である。モンテカルロ法を用いたゲーム木探索手法のうち最も単純なものは、単純なモンテカルロ法と呼ばれる。

ある局面で着手可能な手のうち 1 つを選択し、そこから勝敗が付くまでランダムに手を出し続ける (プレイアウト)。そしてその結果に応じた評価値 (プレイアウト報酬) を、選択した手に与える。これを着手可能な手それぞれで繰り返し行い、最も評価値の高い手を最善手として選択する。

この手法は、手の選択がランダムであるため、明らかな悪手に対してもシミュレーションを行ってしまう問題がある。

3.2 UCT

UCT とは UCB (Upper Confidence Bound) 値を用いて単純なモンテカルロ法の探索効率を上げた手法である。プレイアウトを行う際に UCB 値の最も高い手を選択することで、より有望な手にプレイアウトを割り振ることが出来る。本研究では、UCB 値の計算に UCB1-tuned [4] と呼ばれる手法を用いた。

3.3 モンテカルロ木探索

モンテカルロ木探索は、UCT を改良した手法である。UCT による手の選択を自分の手番以外でも行うことで、最小最大戦略に沿った探索を行うことが出来る。

4 先行研究

福島ら [5] は UNO に適した戦略を見つけるために、平均点数、平均順位、上がる頻度、特定の相手の得点の低さを報酬とした 4 つの単純なモンテカルロ法を用いた

プレイヤーを作成した。また、比較用にヒューリスティックプレイヤー及びランダムプレイヤーを作成した。

対戦の結果、全ての単純なモンテカルロ法を用いたプレイヤーがランダムプレイヤーより強く、平均順位を報酬としたプレイヤーがヒューリスティックプレイヤーに対して優位であることを示した。

5 実験方法

対戦実験を行うにあたって以下のプレイヤーを実装した。

R ランダムに手を選択するプレイヤー

H 自身の上がりを目指しつつ、失点を抑える戦略を実装したヒューリスティックプレイヤー

MCS 単純なモンテカルロ法を用いたプレイヤー

UCT 単純なモンテカルロ法の手選択に、UCB 値を用いたプレイヤー

MCTS モンテカルロ木探索を用いたプレイヤー

MCS, UCT, MCTS のプレイアウト報酬はいずれも平均順位とし、プレイアウト回数は各手に対して 2000 回とした。

上記プレイヤーのうち R と H を性能比較用プレイヤーとし、MCS, UCT, MCTS それぞれと対戦を行った。

6 実験結果

まず R 3 体と、H, MCS, UCT, MCTS それぞれとの対戦を 10000 回行った。理論上は同一プレイヤー 4 体で対戦を行えば、いずれのプレイヤーも勝率は 0.25、平均得点は 0 となる。対戦結果を表 1 に示す。

表 1: R 3 体との対戦結果

	H	MCS	UCT	MCTS
勝率	0.30	0.31	0.31	0.24
平均得点	12.57	15.40	15.22	-0.60

結果より、H, MCS, UCT は R より優れていた。一方、MCTS は R よりも僅かに劣る性能であった。

次に、R よりも優れていた MCS と UCT それぞれと、H 3 体との対戦を 10000 回行った。対戦結果を表 2 に示す。

結果より、MCS, UCT とともに僅かに H より優れていた。

表 2: H 3 体との対戦結果

	MCS	UCT
勝率	0.26	0.27
平均得点	1.90	3.26

7 考察と今後の課題

囲碁や大貧民で有効とされるモンテカルロ木探索は、UNO においては有効ではなかった。MCTS の性能が MCS や UCT, H よりも著しく低かったことについて考察する。

MCS や UCT では、プレイアウトを行うたびに相手の手札をランダムに割り当てている。しかし MCTS ではゲーム木を構築するために、ランダム割り当てを一度行った後は、全てのプレイアウトで共通の手札を使用している。UNO は総カード数の多さと比較して初めに配られる手札枚数は少ない。またゲームの途中で手札が増えることもあり、相手の手札の予想が非常に難しい。このことから探索精度が低下し、MCTS の性能が最も低い結果になったと考えられる。

今後の課題として、相手手札の推定による探索精度の向上が挙げられる。より正確な手札でプレイアウトを行うことが出来れば性能が向上すると考えられる。

参考文献

- [1] David Silver, Aja Huang, and et al. Chris J. Maddison. Mastering the game of Go with deep neural networks and tree search. *Nature*, Vol. 529, pp. 484–489, 2016.
- [2] 電気通信大学 西野・若月研究室. UECda-2022 コンピュータ大貧民大会. <http://www.tnlab.inf.uec.ac.jp/daihinmin/2022/>. 閲覧日: 2023-01-11.
- [3] プログラミングコンテスト ALGORI. <https://www.d3.ntt-east.co.jp/algori/lp.html>. 閲覧日: 2023-01-11.
- [4] Volodymyr Kuleshov and Doina Precup. Algorithms for the multi-armed bandit problem. *arXiv preprint arXiv:1402.6028*, 2014.
- [5] 福島健介, 鈴木徹也. 不完全情報ゲーム UNO のモンテカルロ法によるプレイヤーの構築. 情報処理学会第 83 回全国大会講演論文集, pp. 319–320, 2021.