

TUBSTAPにおけるユニット別攻撃行動枝刈りの効果

提橋凜^{1,a)} 西野順二²

概要: 複数着手性があり分岐因子の大きなターン制戦略ゲームプレイヤーにおける枝刈りの効果を、共通プラットフォームである TUBSTAP を対象に検討した。ターン制戦略ゲームは1ターンあたりの分岐因子が億を超え実質的に2ターン以上の完全な探索が不可能であり、モンテカルロ探索においても、効果的な枝刈りが必須である。本研究では、特にユニットごとの攻撃行動にもとづいて枝刈りを行う手法を検討し、攻撃行動を代表的なものだけに削減するプレイヤーについて、対戦実験による評価を行った。実験した条件の範囲では、結果として、勝率をあげる効果は見られなかったことから、攻撃効果の大小よりも、事後のユニットの陣形のほうが重要であることが示された。

Attack Action Pruning in Unit based UCT for TUBSTAP

RIN SAGEHASHI^{1,a)} JUNJI NISHINO²

Abstract: We examined the effect of pruning in multiple turnbased strategy game that have multiple unit movement in turn. Experiences is done with TUBSTAP which is a common platform. Turnbased strategy game has more than 100 million branching factors per turn. It is practically impossible to search a complete two or more turns. Therefore, effective pruning is indispensable also in the Monte Carlo search. In this research, in particular, we studied a method of pruning based on unit attack behavior. For experiments, a player that reduces attack behavior to only representative one was used. As a result, within the range of the experimented conditions, there was no effect of increasing the win rate. From this it was shown that the formation of the posterior unit is more important than the magnitude of the attack effect.

1. はじめに

本研究は複数着手性があり分岐因子の大きなターン制戦略ゲームプレイヤーにおける枝刈りの効果を、共通プラットフォームである TUBSTAP [1] を対象に検討する。

近年、囲碁や将棋など困難と思われていた古典的な二人零和完全情報ゲームについて、機械学習による評価値同定と大規模なニューラルネットによる候補手生成によって大きな進歩が得られた。どちらもその発展の基礎として数万を超える大規模な人間プレイヤーの良質な棋譜の存在があげられる。また分岐因子の大きい囲碁においてはモンテカル

ロ木探索の効果が大きい。これらのゲームでは、かつてはヒューリスティックな枝刈り手法が検討されることが多かったが、最近では重み付評価や証明数などを基準に広く深い探索が主として研究されている。

一方、ターン制戦略ゲームはゲームがポピュラーでなく囲碁将棋のようなプロプレイヤーも存在しないため、人間のプレイヤーによる棋譜がほとんど存在しない。また、マップの自由度が高くユニットや盤面の評価も統一的でないため棋譜をそれぞれのマップごとに収集しなければならず数万単位での棋譜データベースの取得はほぼ不可能である。1ターンあたりの分岐因子が億を超え実質的に2ターン以上の完全な探索が実質的に不可能である。この大規模なゲームに対して、人プレイヤーの思考は大胆な着手の刈り込みを行いながら有効な手を発見している。このため、ヒューリスティックな枝刈りがいまだ効果的であると考えられる余地がある。

¹ 電気通信大学 情報理工学研究所
Graduate School of Informatics and Engineering, The University of
Electro-Communications

〒182-8585 東京都調布市調布ヶ丘1丁目5-1

² 電気通信大学

The University of Electro-Communications

a) s1731075@edu.cc.uec.ac.jp

本研究では、大会で顕著な成績を上げている M-UCT [2] に枝刈りを加えその効果を測る。特に、ユニットごとの攻撃行動にもとづいて枝刈りを行う手法を検討する。

TUBSTAP を対象として、提案されている枝刈りの手法として深さ限定モンテカルロ法 [3] がある。1 ターンの合法手の全体、すなわち全てのユニットの全行動の組みから、100 程度の設定数だけ候補生成し、それぞれについて深さを限定した単純モンテカルロ法により評価を行い勝率の高い着手を選択する手法であるが、プレイヤーとしての強さは高くなかった。

2. ユニット別攻撃行動枝刈り

ユニット別攻撃行動枝刈りは、ある局面での特定のユニットの攻撃行動を指定数に削減する枝刈りである。人工プレイヤーより人の方が TUBSTAP で強いのが現状である。この人の行動を内省すると、まず強い攻撃の可能性についての絞り込みを行い、そのサポートとして他のユニットの行動を擦り合わせることもある。ユニット別攻撃行動枝刈りは、このような人の思考のモデル化を発端にしている。

攻撃行動は必ず反撃を受けるため、自軍の合計 HP は単調に減少する。人の思考ではそもそも攻撃をすべきかどうかは、最も攻撃効果が高い方法を検討して決めることが多い。攻撃行動が複数可能であるとき、攻撃効果がより少ないものを選択する理由はほとんどないためである。

また、ターン制戦略ゲームでのユニットの使用方法として、主として攻撃を担う場合と、壁として相手の進路を制限する場合がある。たとえば TUBSTAP の場合、戦車は主に攻撃に使い、歩兵は壁として使用する。このとき攻撃行動に求められる目的がユニットによって大きく異なるため、攻撃行動の枝刈りをユニットごとに行うことの効果が高いと期待した。

2.1 攻撃行動枝刈りプレイヤー

TUBSTAP の標準実装プレイヤーである M-UCT [2] を改造し、特定の種類のユニットの攻撃行動を枝刈りするコンピュータプレイヤーを作製した。M-UCT はユニットの行動（一回の着手）をノードとした木（ユニット行動木）で UCT 探索を行うコンピュータプレイヤーである。

2.2 攻撃行動評価値と枝刈り

行動の枝刈りには、村山ら [4] が提案した攻撃行動の評価法を用いた。攻撃行動の評価値 v は、攻撃で与えたダメージを D_a 、反撃でうけたダメージを D_r 、ダメージを与えた敵ユニットの価値を v_o 、現在の自ユニットの価値を v_s 、とすると、

$$v = (D_a \times v_o) - (D_r \times v_s) \quad (1)$$

である。

ただし、敵ユニットの価値は

$v_o = 10 +$ 行動済みユニットへの予測最大ダメージであり、自ユニットの価値は $v_s = 10 +$ 残り HP である。

本手法では、子ノードの展開は親ノードの局面において未行動のユニットの行動を子ノード集合に追加することで行う。このとき、特定の種類のユニットの攻撃行動については評価を行い、評価値がしきい値以上だった攻撃行動の中で、評価値が高い攻撃行動のみを子ノード集合に追加する。特定の種類以外のユニットの全攻撃行動と全てのユニットの全移動行動は無条件に子ノード集合に追加する。アルゴリズム 1 に子ノード展開と枝刈りの動作を示す。

アルゴリズム 1: 子ノードの展開と枝刈り

```

1: function 子ノード展開
2:   for ユニット  $u$  in 未行動ユニット集合 do
3:     if  $u$  は枝刈りの対象 then
4:        $A_{\text{atk}}(s, u) \leftarrow$  局面  $s$  の  $u$  の攻撃行動集合
5:        $A_{\text{atk}}(s, u)$  から評価値がしきい値未満の要素を除外
6:        $A_{\text{atk}}(s, u)$  を要素の評価値でソート
7:       子ノード集合に  $A_{\text{atk}}(s, u)$  から評価値が高い順に指定した数まで追加
8:       子ノード集合に  $u$  の全移動行動を追加
9:     else
10:      子ノード集合に  $u$  の全行動を追加

```

3. M-UCT

M-UCT は武藤らによって提案された、ユニット行動木 UCT 探索を用いる TUBSTAP のコンピュータプレイヤーである。本研究では、M-UCT [2] に提案する枝刈りを加えて実験を行った。

M-UCT のアルゴリズムを、アルゴリズム 2 に示す。

3.1 ユニット行動木

ターン制戦略ゲームで UCT を行うとき 1 ノードあたりの分岐数、すなわち合法手の多さが課題となる。

ユニット行動木は、ユニットの行動 1 つをノードに割り当てた探索木である。1 ターンに複数のユニットを行動させることができる場合、ユニットの数と等しい深さまで探索することで 1 ターンの行動を得ることができる。ユニット行動木において、1 つのノードが持つ子ノードの数は高々各ユニットが可能な行動の総和であり、現実的な時間で全ての子ノードを評価することが可能である。

3.2 UCT 探索

M-UCT はユニットごとに生成する木に対して、UCT(Upper Confidence Bounds for Trees) 探索を行う。UCT はノードの有望さの指標に UCB 値を用いるモンテカルロ木探索である。ノード i の UCB 値はプレイアウトの報酬の期待値 \bar{x}_i 、ノード i を選択してプレイアウトを行った回数

n_i , 探索中にプレイアウトが行われた総数 n , UCB 値の性質を示す定数 c を用いて

$$UCB = \bar{X}_i + c\sqrt{\frac{\log n}{n_i}} \quad (2)$$

と定義される.

アルゴリズム 2 : M-UCT

```

1: function M-UCT
2:    $s \leftarrow$  根ノードの局面
3:   :LOOP
4:    $A(s) \leftarrow$  未行動ユニットの行動の集合
5:    $a^* \leftarrow \operatorname{argmax}_{a \in A(s)}$   $a$  の UCB 値
6:    $s'(s, a^*) \leftarrow s$  で  $a^*$  を実行した後の局面
7:    $s \leftarrow s'(s, a^*)$ 
8:   if  $s$  は葉ノードではない then
9:     GOTO LOOP
10:  if  $s$  のプレイアウト回数  $\geq$  しきい値 then
11:     $s$  の子ノードを展開
12:    GOTO LOOP
13:  else
14:     $s$  のプレイアウトを行う
15:    プレイアウトの結果を辿ってきたノードに反映

```

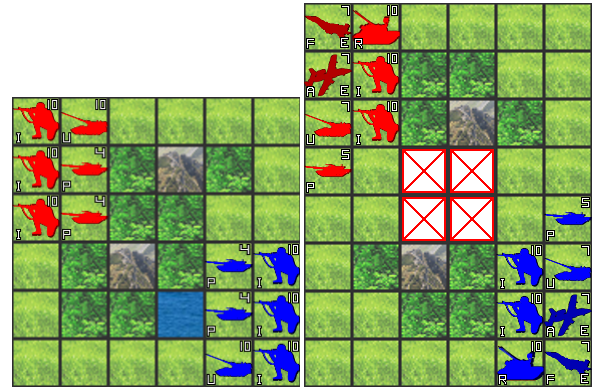


図 1 実験マップ 1

図 2 実験マップ 2

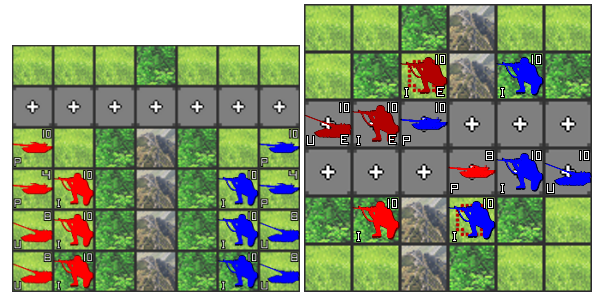


図 3 実験マップ 3

図 4 実験マップ 4

4. 対戦実験による枝刈りの効果の測定

M-UCT をベースラインプレイヤーとして対戦を行い, 枝刈りの対象とするユニットの種類による勝率への効果を実験的に比較する.

4.1 探索の設定

提案手法, M-UCT とともに, UCT 探索に関わるパラメータは同じ値を使用し, プレイアウトのシミュレーション方策も同一とした.

定数 c は 0.15, 子ノード展開のしきい値は 10 とした. プレイアウトの報酬は, 結果が勝利のとき 1, 敗北のとき 0, 引分のとき 0.5 とした. 報酬の期待値 \bar{X}_i の値域は $[0, 1]$ である.

1 ターンの持ち時間は 5000 ミリ秒とし, 1 ユニットの行動決定に使用する時間は残りの持ち時間を未行動ユニットの数で等分したものとした.

4.2 使用マップ

実験には図 1 の実験マップ 1, 図 2 の実験マップ 2, 図 3 の実験マップ 3, 図 4 の実験マップ 4 を使用した. いずれも TUBSTAP に標準実装されているマップである.

4.3 実験 1 単一選択枝刈り

子ノード集合の中から, 最も評価値が高い攻撃行動以外を削除する枝刈りを行った.

実験マップ 1,2,3,4 で行った実験の結果を表 1 に示す.

表 1 実験 1 の結果

マップ	枝刈対象	手番	回数	勝利	敗北	引分
実験マップ 1	歩兵	先攻	550	253	272	25
		後攻	550	181	350	19
	戦車	先攻	1650	801	783	66
		後攻	1650	389	1194	67
	自走砲	先攻	350	166	170	14
		後攻	350	101	236	13
実験マップ 2	歩兵	先攻	550	203	251	96
		後攻	550	221	215	114
	戦車	先攻	250	67	134	49
		後攻	250	89	97	64
	戦闘機	先攻	250	78	123	49
		後攻	250	106	92	52
実験マップ 3	歩兵	先攻	500	268	165	67
		後攻	500	125	314	61
	戦車	先攻	1000	468	400	132
		後攻	1000	186	686	128
実験マップ 4	戦車	先攻	1000	17	956	27
		後攻	1000	709	138	153

4.4 実験 2 半数枝刈

子ノード集合の中の攻撃行動のなかから, 評価値が低い順に半数を削除する枝刈りを行った.

実験マップ 1,3 で行った実験の結果を表 2 に示す.

4.5 枝刈りの効果

攻撃行動について最大評価値の行動一つにする枝刈りの結果 表 1 と, 可能な攻撃行動を半数に削減する枝刈りのそ

表2 実験2の結果

マップ	枝刈対象	手番	回数	勝利	敗北	引分
実験マップ1	歩兵	先攻	150	69	74	7
		後攻	150	39	103	8
	戦車	先攻	150	66	76	8
		後攻	150	39	106	5
実験マップ3	歩兵	先攻	150	80	49	21
		後攻	150	40	90	20

それぞれの結果表2では、ベースラインプレイヤーのM-UCTに負け越しており、どちらの手法でも勝率を向上させる効果は見られなかった

M-UCTに比べて、枝刈りを行うことで弱くなった理由としては、ヒューリスティックな枝刈りと同じく、主に本来有望であったノードを枝刈りしてしまったことが大きいと考えられる。とくに実験2では、上位半数の候補手を残しているにもかかわらず、やはり勝率が下がっている。

よって、有効な手が攻撃効果の少ない攻撃行動にもある、ということが明らかとなった。攻撃という接近行動自体が重要であり、必ずしもダメージだけでは効果が測れないと言える。このことは、隊列の位置取り、陣形が重要であることを示唆している。

また、枝刈りを行うことでその他の選択肢をより深く探索することを期待したが、これも勝率への影響が少なかったといえる。

5. 攻撃行動枝刈りの探索深さへの効果

枝刈りの探索深さへの効果を調べるために、枝刈りを行う提案手法と枝刈りを行わないベースプレイヤーの同一局面における探索に着目し、実験を行った。実験に用いたのは図1に示されている実験マップ1の初期局面の赤の手番である。枝刈りを行わないベースプレイヤーと戦車を対象に単一選択枝刈りを行うプレイヤーでプレイアウトの回数が20000回に達するまでの探索をそれぞれ100回行い、各深さのノードの個数の平均値を表3に、探索木の深さの平均値を表4に示す。

5.1 探索深さへの効果

表3,表4より、実験した条件の範囲では、攻撃行動枝刈りによる、探索深さへの効果は見られなかった。

6. おわりに

本研究では、ターン制戦略ゲームの研究基盤プロジェクトが提供するTUBSTAPにおいて、ユニット行動木UCT探索にユニットの種類ごとの攻撃行動の枝刈りを加え、その効果を検討した。TUBSTAP標準実装プレイヤーであるM-UCTを改造し、子ノード集合から最も評価値が高い攻撃行動以外を削除する単一選択枝刈り、子ノード集合から評価値が低い順に半数を削除する半数枝刈りを実装し、M-UCTとの

表3 各深さのノード数の平均値

深さ	枝刈無し 単一選択(戦車)	
	枝刈無し	単一選択(戦車)
0	1	1
1	73	71
2	1421.5	1352.2
3	2041.6	2132.33
4	1727.1	1993.1
5	957.4	954.4
6	718.7	657.4
7	12548.1	10893.6
8	2164.9	2095.6
9	1459.2	1081.8
10	531.5	477.9
11	106.9	102.87
12	2.99	6.66
13	0	23.33

表4 探索木の深さの平均値

プレイヤー	平均深さ
枝刈りなし	8.63
戦車	8.73

対戦実験による評価を行ったが、勝率を向上させるような効果は見られなかった。この結果から、攻撃効果の低い攻撃行動にも有効な行動が含まれていると考えられ、攻撃行動評価よりも、陣形や隊列の構築が重要であることが示された。

今後は陣形や隊列などに考慮した評価や探索などを検討していきたい。

参考文献

- [1] ターン制戦略ゲーム学術用基盤プロジェクト, <http://www.jaist.ac.jp/is/labs/ikeda-lab/tbs/> (2017年10月13日アクセス)
- [2] Fuzzy Evaluation of Macroscopic Situation for Turn Based Strategic Games, Kosuke Muto, Junji Nishino, 2016 Joint 8th International Conference on Soft Computing and Intelligent Systems (SCIS) and 17th International Symposium on Advanced Intelligent Systems (ISIS), 2016.
- [3] ターン制ストラテジーのための状態評価型深さ限定モンテカルロ法における消極的行動の抑制, 藤木翼, 村山公志朗, 池田心, ゲームプログラミングワークショップ2014論文集, pp.32-39, 2014.
- [4] 学術研究用プラットフォームとしての大戦略系ゲームのルール提案, 村山公志朗, 藤木翼, 池田心, ゲームプログラミングワークショップ2013論文集, pp.146-153, 2013.