

ターン制戦略ゲームにおける合法手数応答型戦略切替の効果

提橋 凜^{1,a)} 武藤 孝輔¹ 西野 順二¹

概要：本研究は、合法手数が多く探索空間が大きなターン制戦略ゲームに対し、局面の複雑さに応じた複数の戦略切替を提案する。ターン制戦略ゲームのプレイヤーアルゴリズムは、局面の状況に応じてその強さが影響を受け、序盤や終盤において得意不得意があった。これまでに提案された手法のうち M-UCT は探索的手法で序盤には弱く、最良優先探索は一手読みであり終盤は弱い、序盤に強い特徴があった。そこで、この二つをサブアルゴリズムとし、局面の複雑さの指標としてユニット数と推定合法手数を用いて序盤と終盤を判別し、着手決定サブアルゴリズムを切り替えるアルゴリズムとした。まず、ゲームの進展と推定合法手数等の変化を調べ、複雑さが序盤に増加しその後減少することを確かめた。次に、サブアルゴリズム切り替えの閾値に対する勝率の変化を実験的に計測したところ、適切な設定により勝率が向上することがわかった。

キーワード：

1. はじめに

近年において、古典的なターン制ゲームのコンピュータプレイヤーの発展は著しく、将棋の ponanza や Google の AlphaGo [1] などが人間のトッププレイヤーを破るといった成果を上げている。これらのゲームのコンピュータプレイヤーの性能は、人間プレイヤーが強さを競い合う目的で使用するに十分な性能を持たせることが出来ているといえる。その一方で、合法手が多いゲームにおいては、探索が困難であることなどの理由から、人間プレイヤーの要求する強さを満たせることは出来ていない。

ターン制戦略ゲームは同一ターンに着手を複数回を行う複数着手性から合法手数が膨大になりやすく、人間プレイヤーと競い合えるコンピュータプレイヤーの作製は困難である。市販のソフトではコンピュータプレイヤー側にハンデを付けることで難易度の調整を行っており、対等な条件で競い合うと行った遊び方は提供されていない [2]。

本研究は、局面の複雑さに応じてアルゴリズムを切り替えるプレイヤーを提案する。ターン制戦略ゲームでは局面の複雑さの変動が大きいため、深い探索を用いる方法と用いない方法をそれぞれが適した局面で切り替えることで、ゲーム全体として有利になることが期待できる。

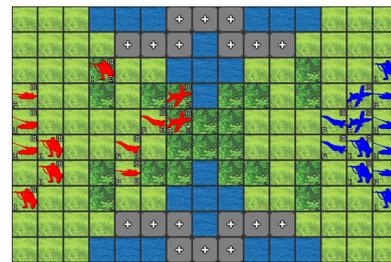


図 1 TUBSTAP のプレイ画面

2. TUBSTAP

TUBSTAP は北陸先端科学技術大学院大学 (JAIST) の村山らが立ち上げたターン制戦略ゲーム学術用基盤プロジェクトにおいて、公開されているターン制戦略ゲームに用いられる思考アルゴリズムの研究・開発のための環境、およびその対象となっているゲームである [3], [4]。プロジェクトが提供している環境には再現性、比較可能性を持つ簡明なルールとグラフィカルユーザーインターフェイスがあらかじめ実装されており、使用者はそれらの実装を行うことなく他手法との比較が行える環境が得られるため、思考アルゴリズムの開発にのみ専念することができる [3], [4]。TUBSTAP のルールは任天堂社のコンピュータゲームである Advance Wars: Days of Ruin [5] のルールを参照して、学術用途に簡略化されており、既存の戦略性ターンゲーム

¹ 電気通信大学
The University of Electro-Communications
a) s1311088@edu.cc.uec.ac.jp

の多くに共通している要素が備わっている [3].

2.1 ゲームのルール

TUBSTAP のルールを説明する. 対象とする TUBSTAP のバージョンは 0.108 [4] である.

2.1.1 ユニット

TUBSTAP では戦闘機 (F), 攻撃機 (A), 戦車 (P), 自走砲 (U), 対空戦車 (R), 歩兵 (I) の 6 種類のユニットが存在する. ユニットは種類毎に移動力と攻撃力の値を持っており, 攻撃による HP の減少量や移動に関して影響を与える.

2.1.2 盤面と地形

TUBSTAP は長方形のマスが敷き詰められた平面の盤面を用いる. 各マスには 1 つの地形が設定されており, TUBSTAP では草原, 森, 山, 海, 道路, 陣地, 進入禁止の 7 種類の地形が存在する. 地形は種類毎に移動コストと地形効果の値を持っており, マスの上を通るユニットの移動やマスの上に居るユニットの受ける HP の減少量に影響を与える. 盤面の大きさや地形の配置, ユニットの初期配置を合わせてマップと呼ぶ.

2.1.3 行動

プレイヤーは 1 ターンに自分チームの全てのユニットを任意の順番で行動をさせることができる. 行動には「移動後に隣接攻撃」, 「移動せず攻撃」, 「移動」, 「移動せず攻撃もしない」の 4 種類がある.

2.1.4 HP

各ユニットは HP の値を持つ. HP は 0 以上 10 以下の整数であり, 攻撃を受けることによって減少する. HP が 0 になったユニットは盤上から除外される.

2.1.5 終了条件

どちらかのチームのユニットが盤上から全て除外されたとき, 盤上に残っているユニットが所属しているチームの勝利である. また, ターン数の上限に達した場合は, 各チームに所属しているユニットの HP の総和の差が引き分け判定のしきい値を超えていれば, HP の総和が高い方のチームの勝利である. HP の総和の差がしきい値以下の場合は引き分けとなる.

3. 既存のアルゴリズム

TUBSTAP に対してはすでに深さ限定モンテカルロなどのアルゴリズムが提案されている. ここではサブアルゴリズムとして使用した最良単独行動と M-UCT について述べる.

3.1 最良単独行動

最良単独行動は村山らによって提案され, TUBSTAP に標準搭載されているアルゴリズムである [3], [4]. 最良単独行動は 1 ユニットの行動を一手とした一手読みを行い, そ

の局面において最も評価値が高い行動を選択する. 合法手数が多い序盤においても性能の低下は小さい.

3.2 M-UCT

M-UCT は武藤らによって提案され, TUBSTAP に標準搭載されているアルゴリズムである [4], [6]. M-UCT はユニットの行動を単位に探索木を展開し, UCT 探索を行う. 合法手数が多い序盤においては性能が低いが, 合法手数の少ない終盤においては高い性能を持つ.

4. 局面複雑度応答戦略切替

本研究では, 局面のユニット数および合法手数に応じて戦略を切り替える手法を提案する.

4.1 局面複雑度

TUBSTAP において, 局面の複雑さそのものを比較することは困難である. そこで, 本研究では複雑さの指標として, ユニット数と合法手数に注目する. ユニット数と合法手数は定量的であり, ゲームの序盤から終盤までに大きく変化する要素であることから, ゲームの序盤と終盤を判別用いることが可能であると考えられる.

4.1.1 ユニット数

局面 s における自分チームのユニットの集合が $U(s)$ のとき, ユニット数は式 (1) に定まる.

$$\text{ユニット数} = |U(s)| \quad (1)$$

4.1.2 推定合法手数

全ての行動の順列をそのターンにおける手としたとき, 複数着手性のあるゲームでは, あるユニットの行動が未行動のユニットの行動の数に影響するため, ターン開始時の局面の状態から合法手数を求めることはできない. そこで, ターン開始時の局面における各ユニットの行動の数がターン終了まで変動しないと仮定した値を推定合法手数とし, 局面の複雑度の指標として用いることとする. 局面 s における自分チームのユニットの集合が $U(s)$, 局面 s におけるユニット u の行動の集合が $A(s, u)$ のとき, 推定合法手数は式 (2) に定まる.

$$\text{推定合法手数} = |U(s)|! \times \prod_{u \in U(s)} |A(s, u)| \quad (2)$$

4.2 戦略の切替

本研究では序盤のサブアルゴリズムとして最良単独行動を, 終盤のサブアルゴリズムとして M-UCT を用いる. ターンの開始時に局面複雑度の値としきい値の比較を行い, 局面複雑度の値がしきい値以上だった場合は最良単独行動を実行し, しきい値未満だった場合は M-UCT を実行する.

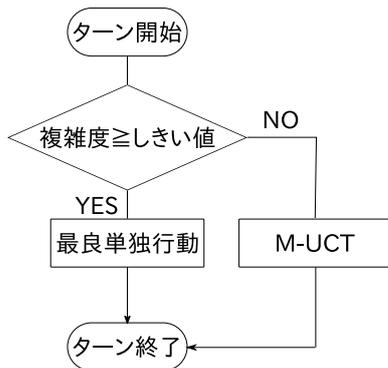


図 2 提案手法のフローチャート

5. 推定合法手数の変化

ゲームの進展と推定合法手数の変化の関係を調べるために、図 3 の実験マップにおける各ターンの推定合法手数の相乗平均値を求めた。

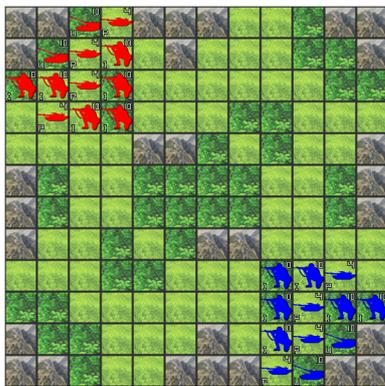


図 3 実験マップ

5.1 実験の結果

実験の結果を図 4 に示す。横軸はターン数であり、縦軸は底を 10 とする平均推定合法手数の対数である。

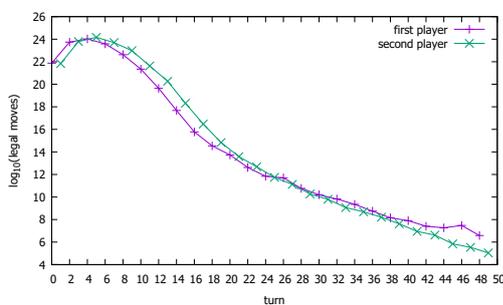


図 4 実験マップ

6. 提案手法の評価

コンピュータプレイヤー同士の対戦実験を行い、提案手法の評価を行う。

6.1 実験の目的

局面の複雑度に応じてアルゴリズムを切り替えるプレイヤーが複雑で大きな局面を含むゲームにおいて有利になることを確認する。また、アルゴリズム切り替えのしきい値と性能の関係を確認すると共に、ユニット数と推定合法手の関係を調べる。

6.2 実験の設定

6.2.1 局面複雑度応答戦略切替の設定

全ての実験において、局面複雑度応答戦略切替で局面が複雑である場合に用いるサブアルゴリズムに最良単独行動、局面が複雑でない場合に用いるサブアルゴリズムにユニット行動木 UCT 探索を用いた。ユニット行動木 UCT 探索の 1 つの行動あたりのプレイアウト回数は 20000 回、ノードのプレイアウト回数のしきい値は 10 回、UCB 値の計算に用いる定数 c の値は 0.15 とした。

複雑度の指標にユニット数を用いる場合は、サブアルゴリズム切り替えのしきい値は 5 から 12 までを 1 ずつ加算し、実験を行った。複雑度の指標に推定合法手数を用いる場合は、サブアルゴリズム切り替えのしきい値は 10^{12} から 10^{23} までを 10 ずつ乗算し、実験を行った。

アルゴリズム切り替えのしきい値を複雑度の値が下回った際のアルゴリズムは比較に用いた M-UCT と全く同一である。

6.2.2 比較に用いたコンピュータプレイヤー

全ての実験において、ユニット行動木 UCT 探索を用いている M-UCT を対戦相手とした。ユニット行動木 UCT 探索の 1 つの行動あたりのプレイアウト回数は 20000 回、ノードのプレイアウト回数のしきい値は 10 回、UCB 値の計算に用いる定数 c の値は 0.15 とした。

6.2.3 使用したマップ

全ての実験において用いた実験マップを図 3 に示す。実験マップのターン数の上限は 50、引き分け判定のしきい値は 10 で赤チームのターンから開始である。

6.3 実験の結果

6.3.1 複雑度の指標にユニット数を用いた場合の結果

複雑度の指標としてユニット数を用いた場合の実験のしきい値に対する勝率の変化を図 5 に示す。勝率は引分を 0.5 勝として計算を行った。

6.3.2 複雑度の指標に推定合法手数を用いた場合の結果

複雑度の指標として s 推定合法手数を用いた場合の実験のしきい値に対する勝率の変化を図 6 に示す。勝率は引分を 0.5 勝として計算を行った。

6.3.3 ユニット数と推定合法手数の散布図

ユニット数と底を 10 とする推定合法手の対数の散布図を図 7 に示す。図 7 の横軸はユニット数、縦軸は底を 10 とする推定合法手数の対数である。

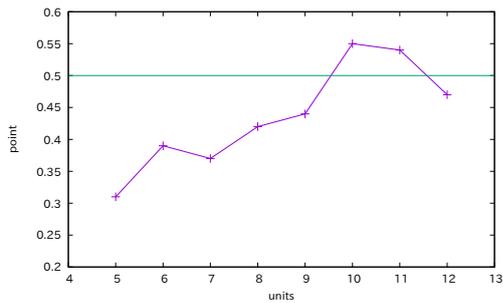


図 5 複雑度にユニット数を用いた場合の勝率

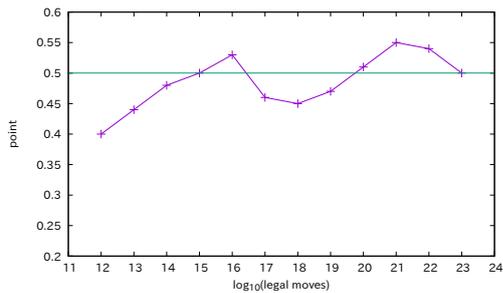


図 6 複雑度に推定合法手数を用いた場合の勝率

6.4 実験の考察

6.4.1 局面複雑度応答戦略切替の有効性の検討

勝利と敗北の差に注目したとき、ユニット数を指標にした際のしきい値が 10,11 の場合、推定合法手を指標にした際の \log_{10} (しきい値) が 16.0,21.0,22.0 の場合について、局面複雑度応答戦略切替は大きく勝ち越していると言える。このことから局面の複雑度に応じて戦略を切り替えることでゲーム全体を有利に運べる場合があることが確かめられた。

6.4.2 アルゴリズム切替のしきい値と性能の関係

6.3 節よりアルゴリズム切替のしきい値によって、勝利及び敗北の数には大きく違いがあると言える。このことから、しきい値の大きさによって局面複雑度応答戦略切替の性能は変動すると考えられる。

また、しきい値の違いは小さいにも関わらず、勝利や敗北の数の差が大きい場合などがあり、しきい値の大きさと性能は非線形だと考えられる。

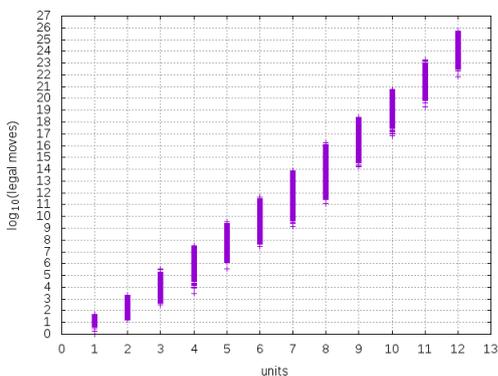


図 7 ユニット数 (横軸) と \log_{10} 推定合法手(縦軸) の散布図

6.4.3 ユニット数と推定合法手数の相関

図 7 よりユニット数と底を 10 とする推定合法手数の対数に正の相関があると考えられる。

7. まとめ

本研究ではターン制戦略ゲーム TUBSTAP を題材に、合法手数やユニット数といった局面の複雑度に応じて複数の着手決定のサブアルゴリズムを切り替える手法である局面複雑度応答戦略切替を提案し、既存手法との対戦実験による評価・検討を行った。その結果、着手決定サブアルゴリズム切り替えのしきい値となる局面の複雑度の値を適切に定めることで、ゲーム全体として有利になることの確認に成功した。

しかし、本研究のようにサブアルゴリズム切り替えのしきい値に対して網羅的に実験を行うことで良い数値を得ることは確実ではあるが、新規のマップに対応させる毎に実験を行うことは実際的ではないため、より簡便に適切なしきい値を推定する方法を発見する必要がある。

本研究の一部は科学研究費補助金基盤研究 C(16K00503)により支援されたものである。

参考文献

- [1] Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M. et al.: Mastering the game of Go with deep neural networks and tree search, *Nature*, Vol. 529, No. 7587, pp. 484–489 (2016).
- [2] 翼 藤木, 公志朗村山, 心 池田: ターン制ストラテジーのための状態評価型深き限定モンテカルロ法における消極的行動の抑制, ゲームプログラミングワークショップ 2014 論文集, Vol. 2014, pp. 32–39 (2014).
- [3] 村山公志朗, 藤木翼, 池田心ほか: 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, 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).
- [4] 池田研究室: ターン制戦略ゲーム学術用基盤プロジェクト, 北陸先端科学技術大学院大学 (JAIST) 情報科学研究科 (オンライン), 入手先 (<http://www.jaist.ac.jp/is/labs/ikeda-lab/tbs/>) (参照 2017-2-9).
- [5] Nintendo: Advance Wars: Days of Ruin, Nintendo (online), available from (<http://www.nintendo.com/games/detail/nLeg9iJkPgq3fWBcqtPDNWUJ4IvmaQBY>) (accessed 2017-2-9).
- [6] 武藤孝輔, 西野順二ほか: ターン制戦略ゲームにおけるファジィ評価を用いた探索木の枝刈り, ゲームプログラミングワークショップ 2015 論文集, Vol. 2015, pp. 54–60 (2015).