2L1-2

ゲーム木探索における満足化の効果

Effective Satisficing in Game Tree Search

大用 庫智[*1] 高橋 達二[*2]
Oyo, Kuratomo Takahashi, Tatsuji

[*1*2] 東京電機大学
Tokyo Denki University

Monte Carlo tree search methods (especially the UCT algorithm), which are proven effective for game AIs, have come into action for real-time games. In that type of games, it is necessary to quickly find appropriate options under extremely limited thinking time. For this reason, the use of UCT is not always appropriate because it requires many simulations to perform well. In this study, we propose a *satisficing* tree search algorithm that entails searching through available alternatives until an acceptable criterion is met. Implementing satisficing behavior with only a simple value function, we enhance the performance of Monte Carlo tree search methods.

1. はじめに

2006 年に囲碁ゲーム AI 研究を劇的に進展させたことで，モンテカルロ木探索（MCTS）に注目が集まった．それ以降，MCTS は囲碁以外のゲーム（e.g., さめがめ，ハーツ，Lines of Action）やその他の問題（e.g., 制約充足問題，スケジューリング）に応用され，その汎用性の高さが示されている[Browne 12]．近年，MCTS はリアルタイムゲーム（e.g., Ms. Pac-Man）への応用も初められ，非常に広い研究領域で扱われている[Browne 12]．

従来の手法（e.g., minimax 法）に必要不可欠な静的評価関数の制作は膨大な労力を必要としていたが，MCTS はその関数の代わりにランダムなサンプリングから計算する勝率を利用することで各ゲームに実装可能である．これまでは，その勝率の代わりに UCB (upper confidence bound) という価値を行動の価値として用いる UCT (UCB applied to trees)が特に利用されてきた．UCT は長期的な運用が可能であればいつかは最適解に到達するが，初期の振る舞いによる指標のぶれと膨大な試行回数が問題になっていた．近年，その代替案として人間の認知の偏りを実装した緩い対称性（LS）モデル[大用 15]という価値関数を行動の価値として用いる LST (LS model applied to trees)が提案されている[Oyo 14]．LST は，人間認知に観られる「受容可能な基準を満たす選択肢の探索」という満足化[Simon 56]を効率的に行う MCTS の一種であり，後に述べる抽象的なゲーム木において満足化基準下であれば UCT よりも高成績を示した．

MCTS は非常に限定された思考時間の中で次の着手を決めなければならないリアルタイムゲーム（例えば Ms. PacMan では 40~60ms）への応用も進んでおり，より計算量が少ない方法が望まれている．そこで本研究では，満足化の単純な価値関数である RS (reference satisficing) モデル[高橋 15]を木探索に応用した RST (RS model applied to trees)を提案する．そして，RST が簡単な価値関数を行動価値としてするだけで，木探索においても満足化の振る舞いが現れ，高成績に繋がることを示す．

2. 抽象的なゲーム木

Kocsis と Szepesvári は抽象的なゲーム木を利用して，UCT が最善手に収束する事を理論とシミュレーションの観点から示した[Kocsis 06]．本研究でも[Kocsis 06, Oyo 14]と同様に図 1 のようなゲーム木を用いる．このゲーム木では，MAX と MIN の二人のプレイヤーが交互にゲームを進める．MAX と MIN は自身にとって最も有望な着手を選択する．ゲーム終局を表す葉ノードには，ゲーム終局の評価値が一定の範囲から一定の確率で与えられる．この評価値が一定の基準よりも高ければ MAX の勝利となり，その基準以下であれば MAX の敗北となる．

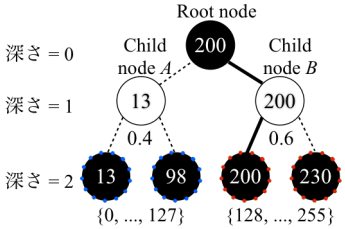

図 1：深さが 2，子ノード数が 2 のゲーム木．赤と青の点線はそれぞれ MAX プレイヤーの勝ち負けを意味している．実線は Mini-Max 的に最適な選択経路である．

二人のプレイヤーの着手の行動価値は葉ノードに到達する事で得られる勝ち（1）負け（0）の情報から計算される．そして，現在の局面（ノード）からその直接の子ノードの行動価値が高い方を選択し，葉ノードまで交互に着手を繰り返す．この行動により勝ち負けの情報をサンプリングする．本来はランダムにゲーム終局まで着手を繰り返すことをプレイアウトと呼ぶが，本研究では上記のサンプリングをプレイアウトと呼ぶ．

2.1 UCT

プレイアウトと実際の着手選択を行うために，UCT ではサンプリングを工夫するアルゴリズム UCB1 [Auer 2002] が利用される．このアルゴリズムはほぼ価値関数にすぎないが，十分な選択回数が許されれば高い成績を示し，期待損失の上界を保証（即ち最適解収束の保証）している．UCB1 アルゴリズムは最初に全ての子ノードを選択しなければならない．その後，価値関数$\mathrm{UCB}(i,j)$の値が最も高い子ノードを選択する．

$$\mathrm{UCB}(i,j) = \bar{X}_{i,j} + \sqrt{2\ln n_i / n_{i,j}}, \qquad (1).$$

ここで，$\bar{X}_{i,j}$は深さ i の子ノード j の期待値（条件付き確率 $P(1|j)$ と一致）であり，$n_{i,j}$は深さ i の子ノード j の選択回数，n_iは（jに限らず）深さ iに到達した選択回数を意味する．

2.2 RST

本研究では価値関数 UCB の代わりに，満足化価値関数である RS モデル [高橋 15] を用いるモンテカルロ木探索を提案する．RST は以下の RS の値が最も高い子ノードを選択する．

連絡先：大用庫智，東京電機大学，kuratomo.oyo[at]gmail.com

$$\mathrm{RS}(1|\mathrm{A}) = (a+d)/(a+b+c+d) \qquad (2).$$

ここで，AとBは図 1 の様な子ノード A と B に対応する．共起情報aは1とAが共に発生した頻度を意味する$N(A,1)$である．同様にb,c,dはそれぞれ$N(A,0)$と$N(B,1)$, $N(B,0)$を意味する．満足化基準 R を用いた RS は$(2\mathrm{R}a + 2\mathrm{R}d) \,/\, (2\mathrm{R}(a+c) + 2\mathrm{R}(b+d))$である．R, $\overline{\mathrm{R}}$ は $\mathrm{R} \in [0,1]$, $\mathrm{R} = 1 - \overline{\mathrm{R}}$を満たす．

3. シミュレーション

ここでは[Kocsis 06, Oyo 14]と同様の設定を用い，最適解収束の速さとその振る舞いの仕方の観点から RST と UCT を比較する．抽象的なゲーム木の深さと一つの親ノードが持つ子ノードの数は，それぞれ 20 と 2 とした．葉ノードの親ノード X には確率P_X が設定されている．葉ノードの評価値が割り振られる範囲は確率P_X で{128, ..., 255}，$1-P_X$ で{0, ..., 127} となる．葉ノードの評価値はその範囲から一様に与えられる．このゲーム終局（葉ノード）の評価値が 128 以上であれば MAX が勝利する．即ち，P_Xが高ければ MAX にとって有利な局面を意味する．この設定に従い 100 種類の木を生成した．そして，それぞれの木に対して 1000 プレイアウトを 100 回実行し，その平均を結果とした．指標は正解率と切り替え率の二つを用いる．正解率は minimax 的に最適なノードを選択した割合である（具体例は図 1 を参照）．切り替え率は前回の選択から選択肢を変えた割合である．これらの指標は各プレイアウト回数の後に計算される．UCB1 には UCB の価値に∞を代入して初期方策を実現する．RS の共起情報には初期値として 1 が代入されている．

3.1 結果

ここでは高確率環境と単高確率環境，低確率環境の三種類の確率設定毎に結果を示す．高確率環境のP_AとP_Bは 0.8 と 0.6 とし，単高確率環境のP_AとP_Bは 0.6 と 0.4，低確率環境のP_AとP_Bは 0.4 と 0.2 とした．つまり，高確率環境では MAX が常に勝利できるような有利な局面を想定した設定であり，低確率環境では常に敗北してしまうような不利な局面を想定した設定である．また単高確率環境では一つの選択肢が唯一勝利可能な局面を想定した設定である．ここで確率の設定は RS のデフォルトである 0.5 を基準としている．

図 2 から UCB1 は環境毎に殆ど切り替え率を変えず，正解率も殆ど同じである．一方，RS は環境毎に切り替え率を変化させており，高確率と単高確率環境では RS の切り替え率は低いが，低確率環境では切り替え率が極端に高くなるという特徴がある．また，RS の正解率は高確率環境では UCB1 に劣るが，単高確率と低確率環境では UCB1 よりも高い事が分かる．図 2 の結果を踏まえると，RS は基準よりも良い局面（高確率と単高確率）では瞬時に勝利する着手に執着する．また，同様に RS は基準よりも低い局面では基準よりも良い局面を探し続けている．つまり，[Oyo 14]と同様な満足化の振る舞いが観られた．

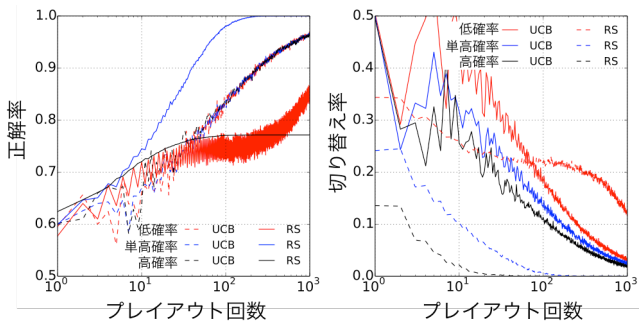

図 2:RS と UCB の正解率（左）と切り替え率（右）

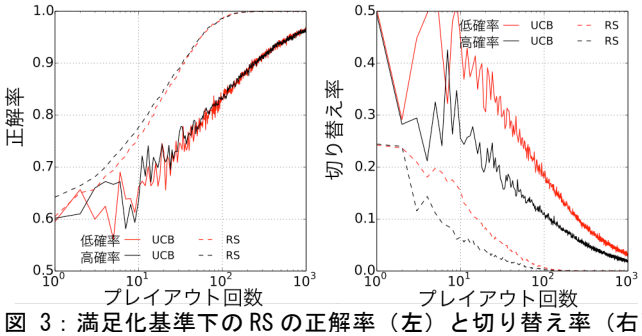

図 3：満足化基準下の RS の正解率（左）と切り替え率（右）

図 2 から RS と UCB1 には振る舞いの仕方に大きな違いがあり，単高確率環境の様に満足化基準が適切であれば UCB1 よりも遥かに高い成績を RS は示す事ができる．そこで，図 2 と同じ設定の高確率と低確率環境において RS の満足化基準を$\min(P_A, P_B) + |P_A - P_B| \times 0.5$と設定した結果を図 3 に示す．

図 3 の結果から RS の正解率と切り替え率が図 2 の単高確率環境と同様になっていることが分かり，UCB1 よりも高成績であることが分かる．これは RS の行った満足化が木探索における最善手の発見という最適化に一致したためと考えられる．

4. 議論と結論

本研究では，囲碁 AI や Ms. Pac-Man で多用されている MCTS に単純な満足化価値関数である RS を実装した RST を提案した．そして，UCT を提案した論文[Kocsis 06]と同様のシミュレーションを通して RST と UCT を比較した．その結果，RST は UCT よりも速く最適解に収束することが分かった（図 2 を参照）．また，RST は木探索においても人間の認知に見られる満足化基準に従い特徴的な振る舞いをすることが分かった（図 3 を参照）．実際のゲーム AI に RST を実装することは有益であると考えられるため，強いゲーム AI またはエンターテイメント用のゲーム AI として RST を具体例に今後実装する．

参考文献

[Auer 02] Auer, P., Cesa-Bianchi, N., and Fischer, P.: Finite-time Analysis of the Multiarmed Bandit Problem, *Machine learning*, 47, 23–256 (2002).

[Oyo 14] Oyo, K., Noguchi, N., and Takahashi, T., Causal Cognition in Game Tree Search, *In Proceedings of 12th International Conference of Numerical Analysis and Applied Mathematics* (ICNAAM 2014) (2014). (in press)

[大用 15] 大用庫智，市野学，高橋達二：緩い対称性を持つ因果的価値関数の認知的妥当性と N 本腕バンディット問題におけるその有効性，人工知能学会論文誌，30(2)，403–416 (2015).

[Simon 56] Simon, H. A.: Rational choice and the structure of the environment, *Psychological Review*, 63(2), 129–138 (1956).

[Kocsis 06] Kocsis, L. and Szepesvári, C.: Bandit based Monte-Carlo Planning, *Machine Learning: ECML 2006 In Proceedings of the 17th European conference on Machine Learning*, 4212, 282–293 (2006).

[高橋 15] 高橋達二，甲野祐，大用庫智，横須賀聡：不確実性の下での満足化を通じた最適化，2015 年度人工知能学会全国大会(第 29 回)予稿集, 2D1-OS-12a-4in (2015).

[Browne 12] Browne, C., Powley, E. Whitehouse, D., Lucas, S., Cowling, P.I., Rohlfshagen, P., Tavener, S., Perez, D., Samothrakis, S., Colton, S.: A survey of Monte Carlo tree search methods, *IEEE Transactions on Computational Intelligence and AI in Games*, 4(1), 1–43 (2012).