タブーサーチを内包したモンテカルロ木探索に基づく囲碁アルゴリズム

太田 雄大[1]　伊藤 雅[2]

概要：モンテカルロ木探索におけるプレイアウトの効率化の研究は活発に行われてきた．しかし，プレイアウトの多様性についての研究はあまりされていない．そこで，本研究ではモダンヒューリスティクスの一つであるタブーサーチをプレイアウトに適用することを提案する．プレイアウトを行った局面をタブーリストに追加し，タブー期間探索するのを禁止する．また，タブー期間を過ぎた局面をタブーリストから取り除く．それによりプレイアウトの多様性を確保することができる．数値実験を対局及び詰碁にて行い，タブーサーチを内包したモンテカルロ木探索は単純なモンテカルロ木探索に比べて良い性能が得られた．

An Igo Algorithm of Monte Carlo Tree Search Including Tabu Search

Takehiro Ohta[1]　Masaru Itoh[2]

Abstract: Efficiency of playout in Monte Carlo tree search (MCTS) have been extensively studied up to now. However, diversity of playout is not really investigated. Because of that, this paper refers to the diversity in MCTS. So this paper proposes to combine MCTS with tabu search (TA), which is a modern heuristic technique for combinatorial problems, into the computer igo algorithm. Once a phase of the playout is added into a tabu list, the searching method prohibits the adoption of the same phase during a given tabu tenure. When the number of trials for playouts is greater than the tabu tenure, the phase is removed from the tabu list. And then the phase could be adopted again. Thus the proposed method can be obtained to ensure the diversity of playout as a whole. The numerical results for some life-and-death igo problems shows that the method of MCTS including TA have obviously got an advantage over the simple MCTS algorithm with the view of the right moves.

1. はじめに

近年のコンピュータ囲碁の棋力は，ここ数年確実に向上している．2013 年 3 月に行われた第 6 回 UEC 杯コンピュータ囲碁大会 [1] では，Crazy Stone[2] が優勝した．

コンピュータ囲碁の棋力が向上した理由として，モンテカルロ木探索 (MCTS: Monte Carlo Tree Search)[3] やプレイアウトの効率化などが挙げられる．関連研究として LGRF(Last Good Reply with Forget)[4] やパターンによるプレイアウトの着手 [5] などがある．前者はプレイアウト中のある局面で勝った打ち方は記憶しておき，負けた打ち方は忘れるという手法である．これにより同じ局面に遭遇した場合，勝った時の手を打つことでプレイアウトを効率化する．後者はパターンをプレイアウトでの候補手とすることでプレイアウトの精度を向上させる．このようにプレイアウトの効率化については活発に研究されているが，プレイアウトの多様性についてはあまり研究されていない．

そこで，本研究ではモダンヒューリスティクスの一つであるタブーサーチ [6] をプレイアウトに適用することを提案する．タブーサーチは，一度探索を行った解をタブーリストに追加し，タブー期間中は探索しないというアルゴリズムである．同一局面の巡回を防ぎつつ，探索空間の多様

1　愛知工業大学大学院経営情報科学研究科
Graduate School of Business Administration and Computer Science, Aichi Institute of Technology
2　愛知工業大学情報科学部情報科学科
Faculty of Information Science, Aichi Institute of Technology

性を確保することができる．

単純なモンテカルロ木探索とタブーサーチを内包したモンテカルロ木探索との性能比較を対局および詰碁にて行った．また，プレイアウトにおける同一局面を探索した比率についても考察する．

2. モンテカルロ木探索

モンテカルロ木探索は，モンテカルロ法に UCB1 アルゴリズム (Upper Confidence Bound)[7] と木探索を用いたアルゴリズムである．単純なモンテカルロ法は，明らかに悪い手にもプレイアウトを均等に実行し，平均勝率で手の良し悪しを判断する．しかし，UCB1 アルゴリズムを使用すれば，勝率が高い手に対してプレイアウトをより多く試行でき，最大 (相手にとっては最小) の価値を持つ手が選択できるようになる．

UCB1 アルゴリズムの計算式を式 (1) に示す．ここで，$\bar{X}_i$ は手 i を打った時の現在の勝率，n_i は手 i 以降のプレイアウトの回数，N はその局面において施行したプレイアウトの総数 (n_i の総和) である．

$$UCB1(i) = \bar{X}_i + \sqrt{\frac{2\log N}{n_i}} \qquad (1)$$

モンテカルロ木探索の処理の流れを以下に示す．

STEP 1 ルートノードから式 (1) の値が最大である子ノードを選択しながら末端ノードまで木を下りる．

STEP 2 末端ノードのプレイアウト回数が閾値を超えていれば，そのノードの子ノードを作成し木を下りる．閾値を超えていなければ，子ノードは作成しない．

STEP 3 末端ノードでプレイアウトを 1 回行う．

STEP 4 辿ってきたノード（ルートノードを含む）すべてにプレイアウトの結果を反映させ，勝率を更新する．

STEP 5 制限時間等の制約に達していれば探索を終了する．さもなければ STEP 1 に戻る．

3. 提案法の概要

囲碁では，図 1 のように着手手順が異なっていても同じ局面になることが多い．特に，9 路盤や詰碁では着手範囲が狭いので顕著に表れる．同一局面の巡回により，局所解に収束する可能性が生じる．同一局面の巡回が起こり易い局面としてシチョウが挙げられる．例えば，シチョウが正着である局面でシチョウを集中的に探索するのは正しいが，シチョウが正着でない局面でシチョウを集中的に探索するのは効果的ではない．

そこで，本研究ではモダンヒューリスティクスの一つであるタブーサーチをモンテカルロ木探索に適用する．

3.1 提案法のアルゴリズム

タブーサーチはモンテカルロ木探索の STEP 3 のプレイアウト中の局面に対して行う．一度探索したプレイアウトの局面をタブーリストに追加しておき，タブー期間中は探索しないようにする．タブーサーチを内包することで，同一局面の巡回を防ぎつつ，プレイアウトの多様性が確保できる．タブーリストは図 2 に示すようにモンテカルロ木の各末端ノードが保持する．図中のノード n1 とノード n2 のタブーリストの要素が異なるように，タブーリストは各ノード毎に管理し，ノード間でのタブーリストの共有は行わない．タブーリストはプレイアウト初手用と次手用の 2 つのリストで構成され，リストの構造はキュー構造とし，リストの長さをタブーサイズとする．図 2 はタブーサイズが 3 であるリストを表しており，図中の First は初手用のリストを，Second は次手用のリストを意味する．また，図 2 中の矢印はタブーリストを更新する際のキューをずらす方向を表している．

前節モンテカルロ木探索 STEP3 への追加アルゴリズムの概要を以下に示す．

STEP 3-1 プレイアウト 1 手目と 2 手目の時，その局面がタブーリストに含まれているか否かを調べる．

STEP 3-2 含まれていなければ，その局面をタブーリストに追加し，プレイアウトを続行する. 含まれているならば，その手は破棄され，プレイアウトを 1 手目からやり直し，STEP 3-1 に戻る．

STEP 3-3 タブーリストに局面を追加すると同時に，タブーリストを更新し，タブー期間を過ぎた局面をタブーリストから取り除く．

タブーサーチを導入することで，単純な MCTS 手法に比べてプレイアウトに要する時間が増大すると推察される．しかし，実際には単純な MCTS 手法と大差のない計算時間であった．次節で提案法の計算量について説明する．

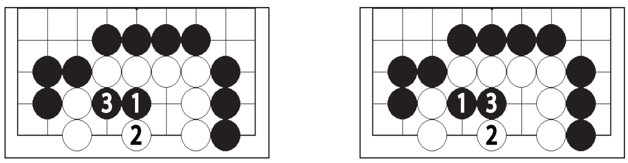

図 1 Example of the same state obtained from different moves

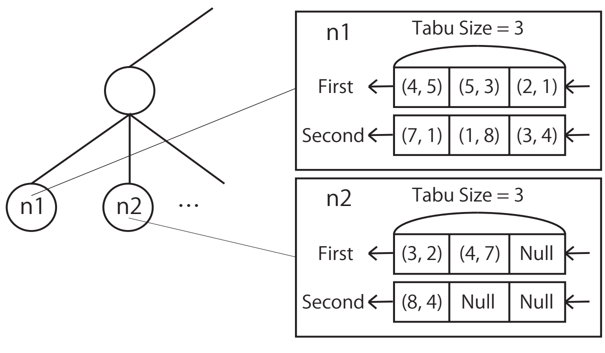

図 2 Tabu lists on each node in the tree

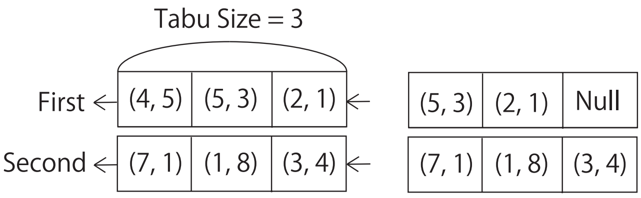

図 3 Updated queue of tabu lists

3.2 提案法の計算量

タブーリストのサイズを L とし，N 路盤で第 K 手目の局面を考える．このときプレイアウトが 1 回成功するまでに実行するプレイアウトの平均試行回数を T とし，T の上界値 $\bar{T}$ を求めてみる．以下の 3 つを仮定する．(1) タブーリストは全部埋まっている．(2) タブーリストに登録する局面は初手と次手の 2 手までとする．(3) 選択された手がタブーリストに含まれる場合，その手は破棄され，タブーリストのキューも同時に更新される．仮定 (3) の状況を図 3 の左図に示す．プレイアウト初手の座標が (5, 3) ならば，初手のリストに (5, 3) が含まれているので，手 (5, 3) は破棄される．その際，初手のリストだけが 1 つずれて，図 3 の右図のようになる．

さて，盤面には石が $(K-1)$ 個あるので，第 K 手目がタブーになる確率は $L/(N^2-(K-1))$ であり，次手がタブーになる確率は $L/(N^2-K)$ である．平均試行回数を上界値で評価すれば，確率 P を $P = L/(N^2-K)$ とおいて Σ(失敗する回数) $\times$ (その確率) $+ 1$ を計算すればよい．

$$\sum_{n=1}^{\infty} n\left\{(n-1)(1-P)P^{n-1} + P^{n}\right\} + 1 \tag{2}$$

式 (2) 第 1 項はタブーサーチを導入したことにより無効となるプレイアウトの平均試行回数の上界値である．プレイアウトが有効になるにはさらにもう 1 回試行を追加しなければならない．それが式 (2) 第 2 項の $+1$ の意味である．式 (2) を計算すると，$3P/(1-P)^2+1$ を得る．上界値という意味合いから，これを $4P/(1-P)^2+1$ で評価すれば，結局 $\bar{T}$ は

$$\bar{T} = \left(\frac{N^2-K+L}{N^2-K-L}\right)^2 \tag{3}$$

として得られる．

式 (3) よりモンテカルロ木探索にタブーサーチを内包させても無効になるプレイアウト数が極端に増えることはないといえる．

4. 数値実験

単純な MCTS 手法とタブーサーチを内包した Tabu + MCTS 手法を対局及び詰碁にて性能比較を行った．また，同一局面の探索比率についても実験を行った．各種パラメータは，タブーリストのサイズを 3〜12，タブー期間を 3〜12，タブーリストに追加する局面をプレイアウト 2 手以内，プレイアウトの閾値を 100 とした．

表 1 Results of the games

結果	対局数	勝ち数	負け数	勝率
Tabu3	200 局	138 勝	62 敗	69.0 %
Tabu4	200 局	148 勝	52 敗	74.0 %
Tabu5	200 局	154 勝	46 敗	77.0 %
Tabu6	200 局	162 勝	38 敗	81.0 %
Tabu7	200 局	152 勝	48 敗	76.0 %
Tabu8	200 局	162 勝	38 敗	81.0 %
Tabu9	200 局	155 勝	45 敗	77.5 %
Tabu10	200 局	155 勝	45 敗	77.5 %
Tabu11	200 局	157 勝	43 敗	78.5 %
Tabu12	200 局	160 勝	40 敗	80.0 %

表 2 Results of binomial test

結果	p 値	信頼区間
Tabu3	$8.027e^{-8}$	0.621 - 0.753
Tabu4	$7.261e^{-12}$	0.673 - 0.799
Tabu5	$1.017e^{-9}$	0.703 - 0.865
Tabu6	$< 2.200e^{-16}$	0.749 - 0.862
Tabu7	$8.738e^{-14}$	0.695 - 0.817
Tabu8	$< 2.200e^{-16}$	0.749 - 0.862
Tabu9	$2.397e^{-15}$	0.711 - 0.831
Tabu10	$2.397e^{-15}$	0.711 - 0.831
Tabu11	$< 2.200e^{-16}$	0.722 - 0.840
Tabu12	$< 2.200e^{-16}$	0.738 - 0.853

4.1 対局による性能比較

対局結果を表 1 に示す．表 1 の Tabu3〜12 はタブーサイズが 3〜12 である Tabu + MCTS 手法を表している．Tabu + MCTS 手法を黒番，単純な MCTS 手法を白番とし，コミを 6 目半，持ち時間を各 10 分，碁盤を 9 路盤とした．

表 1 より，Tabu3〜Tabu12 の全てにおいて単純な MCTS 手法に対して高い勝率が得られた．次に，表 1 の対戦結果を基に有意水準 5%で二項検定を行った．結果を表 2 に示す．ここで，帰無仮説を「単純な MCTS 手法と Tabu3〜Tabu12 の Tabu + MCTS 手法に棋力の差はない」とし，対立仮説を「二つの手法に棋力の差がある」とした．表 2 より Tabu3〜Tabu12 の全てにおいて p 値が 0.05 より小さいのは明らかである．よって，帰無仮説は棄却され，対立仮説が採択される．つまり，Tabu + MCTS 手法と単純な MCTS 手法には棋力の差があるといえる．

4.2 同一局面の探索比率

単純な MCTS 手法と Tabu + MCTS 手法が，それぞれ 9 路盤の黒番 1 手目を探索した際のプレイアウトにおいて同一局面を探索した比率を図 4 に示す．また局面の対象は，プレイアウト 1 手目と 2 手目の局面とした．

図 4 より，Tabu + MCTS 手法の方が単純な MCTS 手法に比べて同一局面を探索した比率が低いことが分かる．

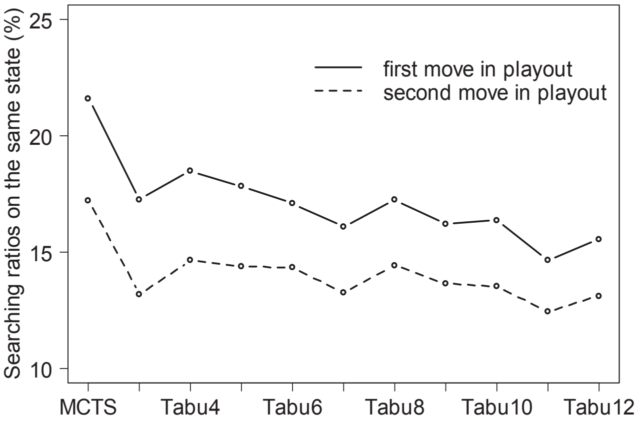

図 4 Searching ratios on the same state in playouts

また，タブーサイズが大きくなるに従って，同一局面を探索した比率が減少している．具体的に，プレイアウト 1 手目で MCTS は 21.5%に対し Tabu11 で 14.6%，プレイアウト 2 手目で MCTS は 17.2%に対し Tabu11 で 12.4%に減少している．これは，提案する Tabu + MCTS 手法が単純な MCTS 手法に比べてプレイアウトの多様性を確保していることを意味する．

4.3 詰碁による性能比較

詰碁 10 問を単純な MCTS 手法と Tabu + MCTS 手法にそれぞれ 30 回ずつ解答させた．それぞれの手法の正答率を図 5 に示す．ここで，正答率を詰碁の 1 手目が正着であった割合とする．図 5 の MCTS とは単純な MCTS 手法を，Tabu12 はタブーサイズ 12 の Tabu + MCTS 手法を表している．

図 5 より MCTS に比べて Tabu12 の方が詰碁正答率が高いことが分かる．同様に Tabu3〜11 に対して同じ実験を行った結果，Tabu12 に近い正答率が得られた．次に，単純な MCTS 手法と Tabu3〜12 の Tabu + MCTS 手法の詰碁正答率の結果を基に有意水準 5%で χ^2 検定を行った．ここで，帰無仮説を「MCTS と Tabu3〜12 に有意な差はない」とし，対立仮説を「MCTS と Tabu3〜12 に有意な差がある」とした．χ^2 検定の結果を表 3 に示す．表 3 より有意水準 5%の χ^2 値である 3.841 を大幅に超えている．よって，帰無仮説は棄却され，対立仮説が採択される．つまり，単純な MCTS 手法と Tabu3〜12 の Tabu + MCTS 手法には有意な差があるといえる．

5. おわりに

本稿ではタブーサーチを内包した MCTS 手法を提案した．数値実験により，対局及び詰碁において提案した Tabu + MCTS 手法の方が単純な MCTS 手法に比べて性能が良いことを確認した．また，Tabu + MCTS 手法はプレイアウトの多様性を確保できることも確認した．

今後の課題として，自己対戦による勝率の分析だけでなくオープンソースとの対戦による性能比較が必要である．

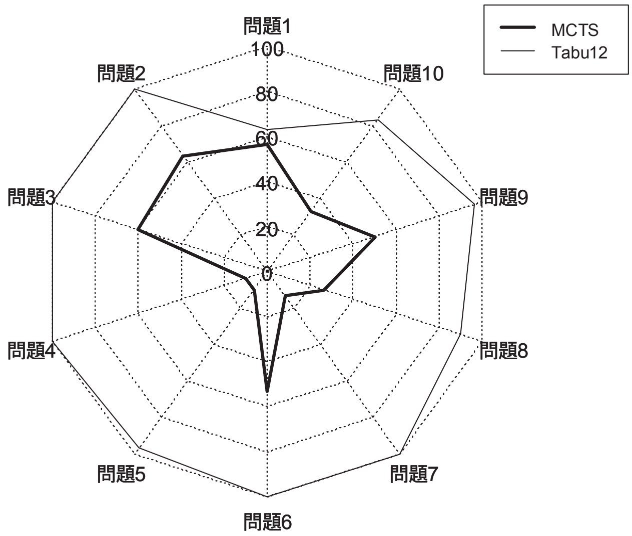

図 5 Ratios of the correct answers

表 3 Results of chi-squared test

結果	χ^2 値	p 値
Tabu3	98.783	$< 2.200e^{-16}$
Tabu4	100.109	$< 2.200e^{-16}$
Tabu5	90.009	$1.622e^{-15}$
Tabu6	99.114	$< 2.200e^{-16}$
Tabu7	95.081	$< 2.200e^{-16}$
Tabu8	107.867	$< 2.200e^{-16}$
Tabu9	103.609	$< 2.200e^{-16}$
Tabu10	83.569	$3.149e^{-14}$
Tabu11	98.841	$< 2.200e^{-16}$
Tabu12	96.830	$< 2.200e^{-16}$

尚，本研究の一部は，文部科学省 科研費 基盤研究 (C) No.25330441 の助成を得た．ここに謝意を表する．

参考文献

[1] 第 6 回 UEC 杯コンピュータ囲碁大会, http://jsb.cs.uec.ac.jp/~igo/

[2] Crazy Stone, http://remi.coulom.free.fr/CrazyStone/

[3] Rémi Coulom, “Computing Elo Ratings of Move Patterns in the Game of Go”, Computer Games Workshop, 2007.

[4] Hendrik Baier, and Peter D. Drake, “The Power of Forgetting: Improving the Last-Good-Reply Policy in Monte Carlo Go”, IEEE Transaction on Computational Intelligence and AI in Games, Vol.2, No.4, pp.303–309, 2010.

[5] Slyvain Gelly, Yizao Wang, Rémi Munos, and Olivier Teytaud, “Modification of UCT with Patterns in Monte-Carlo Go”, Technical Report RR-6062, INRIA, 2006.

[6] 野々部宏司, 柳浦睦憲, “局所探索法とその拡張 —タブー探索法を中心として”, 計測と制御, Vol.47, No.6, pp.493–499, 2008.

[7] 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.