コンピュータ囲碁におけるプレイアウト情報に基づく局面評価を用いたモンテカルロ木探索

田中 一樹[1,a)]　徳重　毅[1]　藤田　玄[1]

概要：
囲碁の局面評価は難しく，コンピュータ囲碁の局面評価関数の多くは，死活判定や探索範囲の広大さなどが原因で評価に時間がかかる．この問題に対し，本研究室では，UCT の中で着手決定に用いられるプレイアウトの終局盤面に注目し，終局状態の局面の統計情報を用い，少ない演算コストで盤面評価を行う手法を提案した．しかしこの手法では子ノード同士の UCB 値が違う場合に提案手法が有効でなくなどの問題があった．そこで本研究では，UCB 値の勝率項にプレイアウトの統計情報に基づく局面評価関数を組み合わせた手法を提案し，木探索の効率化を図る．既存手法である Fuego に提案手法を適用して性能評価を行った．提案手法の有効な範囲を調べるためにプレイアウト回数や持ち時間を変え評価を行った．その結果，少ないプレイアウト回数では，UCT 単独の Fuego に対して 6 割の勝率を示した．

Kazuki Tanaka[1,a)]　Tsuyoshi Tokushige[1]　Gen Fujita[1]

Abstract: In recent years, a Monte Carlo tree search such as UCT for computer go have been widely known. On the other hand, an evaluation functions of the state of the game of computer go still tends to be heavy computational cost. In the previous study, we had proposed an efficient evaluation function which focus on results of play out based on UCT. However, the method had only limited effect, because UCB value was preserved. To cope with this problem, in this paper, an improved UCB value including a statistical information of results of playout. The evaluation results show some effectiveness of the proposed method.

1. はじめに

2000 年代からのコンピュータ囲碁では，評価関数に頼らない，モンテカルロ木探索 [1]，主に探索に期待値 UCB（Upper Confidence Bound）[2] を用いた UCT(UCB applied to Trees)[3] が主流になっている．

UCT の性能を向上させる手法として UCT+[4] がある．この UCT+ は局面評価関数を UCB 値に用いる手法で，オセロにおいて実装・実験され，有効性も実証されている．しかし囲碁では，死活判定の難しさ，探索する範囲が広大，短期的な良手が長期的な良手になるとは限らないなどの理由で途中局面の有効な局面評価を行うことが難しい [5]．それによりコンピュータ囲碁では，モンテカルロ木探索と局面評価関数を組み合わせた研究は少ない．

そこで本研究室では，コンピュータ囲碁で用いる局面評価として，プレイアウトの終局盤面の情報を集計して利用する評価手法 PTS(Playout Territory Statistics) を提案し，局面評価に基づいた探索候補手の効率化の手法 [6][7] に用いた．具体的には，PTS の統計情報の一つであるモンテカルロオーナー (MO)[8] を用いた．しかし MO を用いた探索候補手選出には終局でのダメなどに着手が集中してしまうなどの問題があった．そこで，対局進行度を用いることで選出処理を切り替える手法を提案した [7]．それぞれの手法で，勝率の向上を図ることができたがどの手法も 6 割にとどまっている．

勝率向上に至らなかった原因として，プレイアウト前の探索候補選出に影響を与えるだけでは，子ノード同士の UCB 値が違った場合，提案手法が意味をなさないことが

[1] 大阪電気通信大学

[a)] mi14a005 @ oecu.jp

ある．また木探索全体では，提案手法が呼び出される頻度が少ないなどが原因として考えられる．

そこで本研究では，呼び出し頻度が多く木探索への影響を与えるため，コンピュータ囲碁におけるモンテカルロ木探索の UCB 値の勝率項にプレイアウトの統計情報に基づく局面評価関数を組み合わせた手法を提案する．

既存手法である Fuego[9] に提案手法を適用して性能評価を行った．また局面評価関数の性能評価を行うため Criticality[10] を Fuego に適用したものとの対局を行った．提案手法の有効な範囲を調べるためにプレイアウト回数や持ち時間を変え評価を行った．

2. 従来手法

2.1 UCT

UCT は，2006 年 9 月に Kocsis と Szepesvari によって発表された，モンテカルロ木探索の代表的なアルゴリズムである [5]．UCT では，UCB 値を利用し，以下の (1) から (5) を時間内に繰り返すことで木を作成する．

(1) UCB1 値による局面選択
根節点から，UCB 値の高い子接点を選択しながら末端節点まで枝をたどる．

(2) 局面の展開
末端の節点のプレイアウト回数が閾値を越えていれば，末端節点の局面から合法手を選出し，末端の子節点として節点を展開する．

(3) プレイアウト
(2) で展開した節点の局面からプレイアウトを行う．

(4) UCB1 値の更新
プレイアウトの結果によって得られた評価値を根節点までたどって節点の値を更新する．

(5) 探索続行の判断
制限時間や総プレイアウト数などの制限に達していなければ終了する．そうでなければ (1) に戻る．

囲碁に置き換えた場合，最後的に根節点の子節点から評価値が一番高い手を選択し着手を決定する．UCT の性能を向上させる手法として UCT+[4] がある．この UCT+ は局面評価関数を UCB 値に用いる手法で，オセロに有効性も実証されている．

しかしコンピュータ囲碁においては，そういった研究は少ない．

2.2 Criticality

Criticality[10] は，Coulom によって提案された手法である，勝敗を決定する重要な位置を求める．どちらが先着するかによって勝敗を左右する重要な点を集計していく．Coulom の手法によると点 x における Criticality の値 $c(x)$ は以下の式で求めることができる．

$$c(x) = \frac{v(x)}{N} - (\frac{w(x)}{N} \times \frac{W}{N} + \frac{b(x)}{N} \times \frac{B}{N}) \quad (1)$$

N：プレイアウトの総数
B：黒が勝ったプレイアウト数
W：白が勝ったプレイアウト数
$b(x)$：黒が点 x を取っていたプレイアウト数
$w(x)$：白が点 x を取っていたプレイアウト数
$v(x)$：プレイアウトに勝ったほうが x を取っていた回数

2.3 PTS(Playout Territory Statistics)

PTS[6] はプレイアウトの最終局面の石の配置の統計を取ることで，盤面評価を行う手法である．

統計情報の一つに MO がある．MO により，現局面まだ着手されていない未確定な座標点が，どちらの色に属しているかを計ることができる．

評価値が取る範囲は+1 から-1 の範囲であり，評価値が-1 に近いほど白に属した地，+1 に近いほど黒に属した地である．0 に近いほどどちらにも属していない．座標 p の $MO(p)$ は式 (2), (3) で求められる．

$$P(n_i, p) = \begin{cases} 1(state(p) = \text{黒のとき}) \\ 0(state(p) = \text{空点のとき}) \\ -1(state(p) = \text{白のとき}) \end{cases} \quad (2)$$

$$MO(p) = \Sigma_{i=1}^{n} \frac{P(n_i, p)}{n} \quad (3)$$

n : プレイアウトの総数
n_i: i 番目のプレイアウト
$state(p)$:座標 p の状態
$P(n_i, p)$:i 番目のプレイアウト時の点 p の状態

従来の手法では $MO(p)$ を評価として使っていたがそのまま評価値に用いると終盤には，ダメなどに着手が集中してしまうなどの問題があった．

3. 提案手法

本研究では, コンピュータ囲碁におけるモンテカルロ木探索の UCB 値の勝率項に局面評価関数を組み合わせた手法を提案する．

提案手法は以下の式 (4) を最大化するノードを選択する．

$$\overline{X_p} + wE(p) + c\sqrt{\frac{2\log n}{n_p}} \quad (4)$$

局面評価に用いる評価関数 $E(p)$ の目的として，プレイアウト中での着手頻度が少ない石が将来的に存在する確率の低い点を探索していくことにある．しかし [7] では，$MO(p)$ の値を直接用いた際にダメなどに着手が集中してしまう問題があった．

そこで本研究では，黒が勝ったプレイアウトと白が勝ったプレイアウトで集計を分け，それぞれの差分を取った絶対値を $|MO(p)|$ にかけ合わせることで意味のない着手を

減らすようにしている．

また，新たなプレイアウトの統計を用いる処理として以下の関数を提案する．

- $B\_MO(p)$:点 p における黒が勝ったプレイアウトの MO
- $W\_MO(p)$:点 p における白が勝ったプレイアウトの MO
- $BW\_diff(p)$:点 p における B_MO と W_MO の差を取った絶対値

$$E(p) = (1.0 - |MO(p)|) \times BW\_diff(p) \quad (5)$$

4. 性能評価

提案手法の性能評価行うため，従来手法である UCT のみの Fuego との対局による評価を行った．また今回用いた局面評価関数の有効性を示すため，Criticality と組み合わせた Fuego との対局による評価を行った．

19 路盤で試合を行い性能評価を行った．また有効な重み係数を調べるためそれぞれでも評価を行った．

以下に今回の評価実験を行った条件と環境を示す．

表 1 評価条件

使用プログラム	Fuego Version1.0
プロセッサ	Intel(R)i7-2600 CPU3.4GHz
メモリ	16GB
OS	Windows7 64bit
一手あたりのプレイアウト回数	500,1000,5000
試合数	500,100

4.1 UCT 単独の Fuego との性能評価

UCT のみの Fuego と提案手法の対局による性能評価を行った．w の値を変えたときの Fuego に対する勝率を図 1〜5 に示す．エラーバーは 95%信頼区間を表す．

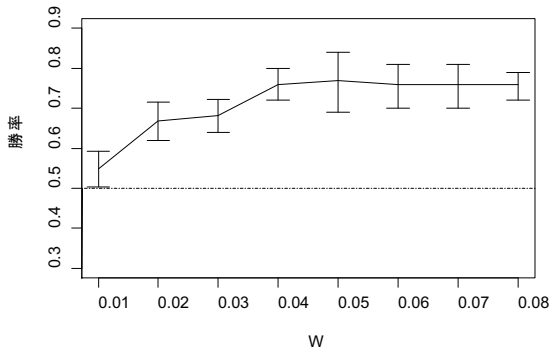

図 1 プレイアウト数 500 回での結果

今回の実験でもっとも提案手法が有効である $w = 0.04$ の結果を以下の表に示す．

以上の結果からプレイアウト数が 500〜1000 までの範囲

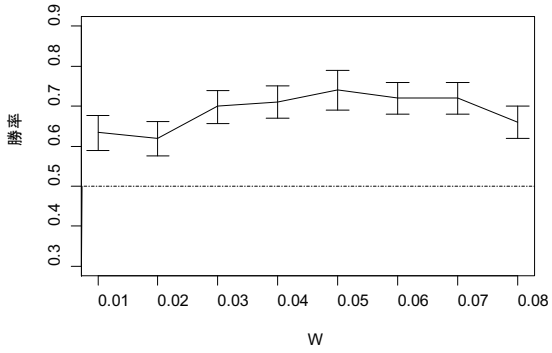

図 2 プレイアウト数 1000 回での結果

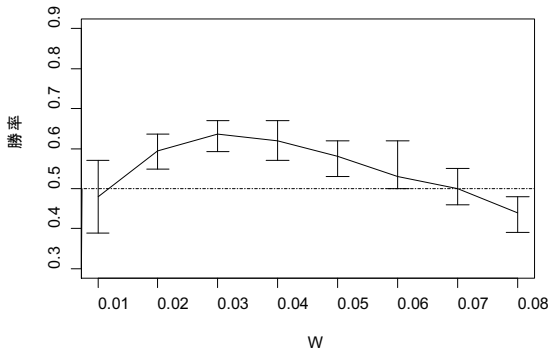

図 3 プレイアウト数 5000 回での結果

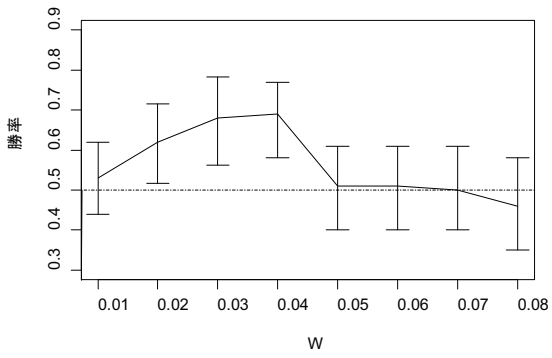

図 4 プレイアウト数 10000 回での結果

表 2 w=0.04 評価結果

playout	500	1000	5000
勝率	76%	71%	62%
信頼区間 (95%)	72%〜80%	67%〜75%	57%〜66%
playout	10000	50000	
勝率	68%	73%	
信頼区間 (95%)	56%〜78%	66%〜79%	

では場合は w の値が有効な範囲が広いが，プレイアウト数が増えていくにつれて w の値が有効な範囲が狭くなってい

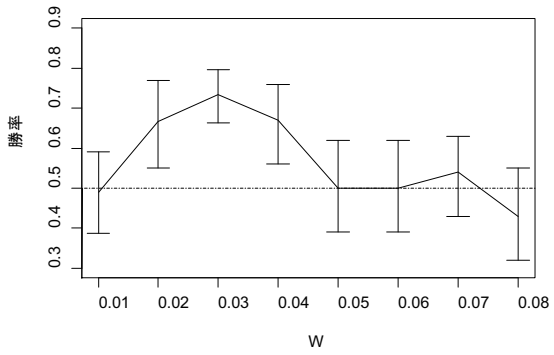

図 5 プレイアウト数 50000 回での結果

ることが分かる．これはプレイアウト数が多くなるにつれて UCB 値の勝率項の信頼度が高まっているからではないかと考えられる．

4.2 Criticality を適用した Fuego との性能評価

提案手法の局面評価の有効性を示すため，局面評価関数 Criticality を UCB 値の勝率項に組み合わせたものとの対局による評価を行った．今回の実験では，Fuego に対して勝率が高った $w = 0.03$〜0.05 を用いる．

今回の実験で用いた式を示す．

$$\overline{X_p} + WC(p) + c\sqrt{\frac{2\log n}{n_p}} \tag{6}$$

また，Criticality に対しても重み付け W を行いそれぞれの値を変えながら評価を行う．

表 3 Criticality と提案手法w= 0.03 の評価結果

W=0.10		
playout	500	1000
勝率	57%	60%
信頼区間 95%	52% ～61%	51% ～69%
W=0.05		
playout	500	1000
勝率	61%	62%
信頼区間 95%	56% ～65%	57% ～66%
W=0.01		
playout	500	1000
勝率	65%	68%
信頼区間 95%	61% ～69%	64% ～72%

表 4 Criticality と提案手法w= 0.04 の評価結果

W=0.10		
playout	500	1000
勝率	59%	59%
信頼区間 95%	54% ～64%	54% ～63%
W=0.05		
playout	500	1000
勝率	68%	59%
信頼区間 95%	64% ～72%	54% ～63%
W=0.01		
playout	500	1000
勝率	71%	69%
信頼区間 95%	67% ～75%	64% ～73%

表 5 Criticality と提案手法w= 0.05 の評価結果

W=0.10		
playout	500	1000
勝率	63%	58%
信頼区間 95%	58% ～67%	54% ～62%
W=0.05		
playout	500	1000
勝率	72%	62%
信頼区間 95%	66% ～78%	58% ～66%
W=0.01		
playout	500	1000
勝率	78%	67%
信頼区間 95%	74% ～82%	53% ～67%

5. まとめ

本研究では, コンピュータ囲碁におけるモンテカルロ木探索の UCB 値の勝率項に局面評価関数を組み合わせた手法を提案した．コンピュータ囲碁におけるプレイアウトの統計情報を用いた局面評価関数を提案した．性能評価の結果からプレイアウト数が大きくなった際に w の値を小さくしていく必要があると考えられる．今後の課題として，プレイアウト数によって動的に w の値を変更するなどが考えられる．

参考文献

[1] Coulom, R.: Efficient selectivity and backup operators in montecarlo tree search, Proceedings of the 5th International Conference on Computers and Games, Turin, Italy (2006).

[2] Auer, P., Cesa-Bianchi, N. and Fischer, P.: Finite time Analysis of the Multi-armed BanditProblem, Machine Learning, Vol. 47, pp. 235-256 (2002).

[3] Kocsis, L. and Szepesvari, C.: Bandied Based Monte-Carlo Planning, Proceedings of the 15th European Conference on Machine Learning, pp.282-293 (2006).

[4] 松本渉, 小林康幸, UCT 探索における局面評価関数の使用方法と性能評価, The 18th Game Programming Workshop 2013

[5] 美添 一樹 , 山下 宏 , コンピュータ囲碁 ―モンテカルロ法の理論と実践―, 共同出版社,(2012).

[6] 田中 一樹，藤田 玄，コンピュータ囲碁における PTS を用いたモンテカルロ木探索，研究報告ゲーム情報学（GI）,2014-GI-32(3),1-4 (2014-06-28)

[7] 田中一樹，藤田玄，コンピュータ囲碁における対局の進行度判定を用いたモンテカルロ木探索，ゲームプログラミングワークショップ 2014 論文集,2014,162-166 (2014-10-31)

[8] R. Coulom. Computing Elo Ratings of Move Patterns in the Game of Go, In Computer Game Workshop,Amsterdam, The Netherlands, 2007.

[9] Fuego，”http://fuego.sourceforge.net/”

[10] R.Coulom. Criticality:a montecarlo heuristic for Go programs. Incited talk at the University of Electro-Communications, Tokyo, Japan,2009.http://remi.coulom.free.fr/Criticality/Criticality.pd