評価関数の違いがモンテカルロ木探索プレイヤの強さに与える影響

1180319　北村 直輝　【 高度プログラミング研究室 】

1　はじめに

ゲーム研究の分野において，探索手法のひとつであるモンテカルロ木探索を改良する手法のひとつとして，評価関数を併用する手法がある．モンテカルロ木探索の乱数部分に対して局面評価によるバイアスをかけることで，短時間でより良い結果を得られると期待される．しかし，評価関数の局面評価の違いなどがモンテカルロ木探索の強さにどのような影響を与えるのかは，著者の知る限り明らかになっていない．

本研究では，既に優れた評価関数が存在するオセロを題材にこの問題に取り組む．評価関数には，オープンソースのオセロプログラム Zebra [1] の評価関数を用いる．また，Zebra の評価関数を改変し異なる評価関数を複数用意し，モンテカルロ木探索でより良い評価関数を用いることの優位性を検証する．

2　モンテカルロ木探索とその改良

モンテカルロ木探索では，ある局面から仮想的に乱数で手を選びながら，終局までプレイするプレイアウトを複数回行い，その結果より手を選ぶ．モンテカルロ木探索では，プレイアウトを有望な手に割り当てることで効率的に探索を行う．さらに，ある局面でのプレイアウト数が増えると，その局面から 1 手進んだ先からプレイアウトを行い，より深く探索することができる．

プレイアウトを無限回行うことで，最適解が得られることが理論的に証明されているが，現実には限られた時間内で解を得ることが求められる．そこで，評価関数による改良を行うことで，短時間でより良い解を得られることが期待される．

改良の例として，プレイアウトの改良と UCB1 値の改良について説明する．

プレイアウトの改良は，プレイアウト中の着手に対し評価関数による評価を行い，局面評価上良いとされる手がプレイアウト中に選ばれやすいようにする．これにより，プレイアウトがある程度強いプレイヤ同士の対戦に近くなる．

UCB1 値は，有望な手の判定に用いられる値である．プレイアウトの勝率が高い，あるいは行われたプレイアウト数が少ない局面の UCB1 値が高くなる．UCB1 値の計算に評価関数を併用し，局面評価上良いとされる手が有望な手と判定されやすくする．

3　評価関数

本研究で用いる Zebra の評価関数は，パターンによる局面評価を行う．盤面から 11 種のパターンを抽出し，その状態に応じて評価値を求める．

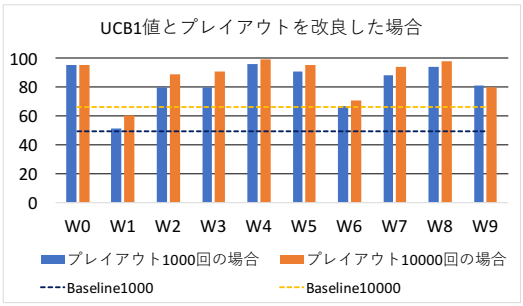

図 1　UCB1 値とプレイアウトを改良した場合の勝数

3.1　評価関数の改変

本研究では，Zebra の評価関数からパターンを 2 つまたは 4 つ除去し改変した評価関数を 9 種作成し，それを W1 から W9 とする．オリジナルを W0 とし，計 10 種の評価関数を用意した．

4　実験

4.1　実験方法

モンテカルロ木探索に，評価関数によるプレイアウトと UCB1 値の改良を適用し，用意した 10 種の評価関数を逐次切り替えて用いる．また，改良も UCB1 値のみ改良した場合，プレイアウトのみ改良した場合，UCB1 値とプレイアウトを改良した場合の 3 通りを用意する．これらのモンテカルロ木探索プレイヤをアルファベータ法プレイヤと対戦させ，その結果を評価する．

4.2　実験結果

実験の結果，評価関数毎に強さに差が現れ，勝数に最大で 4 割程度の差がついた．ただし，改良を行なった場所によってその差が変化し，最大で 2 割程度の差に留まる場合もあった．

図 1 に結果の一部を示す．横軸が評価関数の種類，縦軸が 100 戦中のモンテカルロ木探索の勝利回数である．

5　まとめ

実験結果より，評価関数の違いはモンテカルロ木探索の強さに，大きな影響を与えることが分かった．ただし，評価関数を用いる場所によって，その影響は変わってくる可能性があると考えられる．

参考文献

[1] Gunnar Andersson: Zebra, `http://radagast.se/othello/index.html`.

[2] 北村直輝, 松崎 公紀: 評価関数の違いがモンテカルロ木探索プレイヤの強さに与える影響. 第 59 回プログラミング・シンポジウム予稿集, pp 173-183, 2018.