

将棋における、評価関数を用いたモンテカルロ木探索

竹内 聖悟^{†1} 金子 知適^{†1} 山口 和紀^{†1}

近年、モンテカルロ木探索が改良され、特に囲碁において強いコンピュータプログラムが生み出されている。Lines of Action, Amazons, Arimaa といったゲームにおいて、モンテカルロ木探索へ評価関数を用いる手法が研究されている。本論文では、将棋へと評価関数を用いるモンテカルロ木探索を実装し、モンテカルロ木探索の問題点の解決と性能改善のために、評価値の差、静止探索を利用する手法を提案する。問題集や自己対戦の実験結果から、性能改善を確認し、提案手法の有効性を示した。

Evaluation Function Based Monte Carlo Tree Search in Shogi

SHOGO TAKEUCHI,^{†1} TOMOYUKI KANEKO^{†1}
and KAZUNORI YAMAGUCHI ^{†1}

Recent improvements on Monte Carlo Tree Search(MCTS) have produced strong computer Go programs. Evaluation function based MCTS (EF-based MCTS) has been approached in the game of Lines Of Actions, Amazons, and Arimaa. In this paper, we apply EF-based MCTS to Shogi and present a method for improving the efficiency of EF-Based MCTS in Shogi. Improvement is confirmed by solving problems and by self-play.

1. はじめに

近年、コンピュータ囲碁の分野においてモンテカルロ木探索が多くのプログラムに用いられ、大きな成果を挙げている¹⁰⁾。その後、General Game Playing³⁾ など、他のゲームでも応用が試みられている。ゲームに関する知識はルール以外不要である点がモンテカルロ木探索の特徴として挙げられるが、性能の改善のためにはゲームの知識を利用する手法が研究されている^{2),4)}。

評価関数は、ゲームの知識から局面の形勢判断を行うものである。モンテカルロ木探索を改良する上で、ゲームの知識として評価関数を利用するのは自然なアイデアであり、これまでに、Amazons^{5),7)}, Arimaa⁶⁾, Lines Of Action (LOA)⁸⁾ のゲームで評価関数を利用したモンテカルロ木探索が提案されている。通常のモンテカルロ木探索ではアルファベータ木探索に性能が及んでいなかったが、評価関数の利用により大きく性能を改善され、LOA ではアルファベータ木探索のプログラムに対し勝率 46%、Amazons では従来のプログラムに対し勝率 80% と、それぞれ同等かそれ以上の

結果を得ている。

将棋へとモンテカルロ木探索を応用することの意義として、モンテカルロ木探索の並列化がアルファベータ木探索に比べ容易であることや、局面によっては、アルファベータ木探索よりも正確な評価を行うことが可能であることなどが挙げられる。また、将棋では棋譜を用いた評価関数の学習が行われており^{12),15)}、評価関数の質が高くなっていると期待されることも理由の1つである。一方、モンテカルロ木探索における問題点として、駒の取り合いのような探索が必要となる局面は不得意であること、優勢な局面においては、どの手を選んででも優勢であるため、良くない手を選択する傾向があることが挙げられる。特に前者は、将棋においては駒の取り合いは多く発生する上に勝敗に直結しやすいため、大きな問題点となる。

本研究では、評価関数を利用したモンテカルロ木探索を将棋へと応用し、モンテカルロ木探索における上記の問題点を解決する手法を提案する。

2. 先行研究

評価関数を利用したモンテカルロ木探索と将棋におけるモンテカルロ木探索について述べる。

モンテカルロ木探索のプレイアウトにおいて評価関数を利用する手法には、一定の深さに達した時に評価

^{†1} 東京大学大学院総合文化研究科

Department of General Systems Studies, Graduate School of Arts and Sciences, The University of Tokyo
{takeuchi,kaneko,yamaguchi}@graco.c.u-tokyo.ac.jp

関数の値から勝敗を定める方法 (Amazons⁷⁾, Arimaa⁶⁾) と、深さに関係なく評価値が閾値を超えた時に勝敗を返す方法 (LOA⁸⁾) がある。前者は Amazons においてアルファベータ木探索のプログラムに対し勝率 80%、後者は LOA においてアルファベータ木探索のプログラムに対し勝率 46% となり、有効なモンテカルロ木探索プログラムを作ること成功している。また、プレイアウト中の指し手の選択に評価値を利用する手法⁸⁾も提案されており、これは、前の局面よりも評価値が悪くなる手は選ばない手法と評価値が最も高い手を選ぶ手法と両者を組み合わせている。なお、本研究で取り扱うプログラムでは、1 手先の評価値を得るのにかかるコストが高く、プレイアウトに時間がかかるため、これらの手法は利用していない。

モンテカルロ木探索の将棋への応用には佐藤らの研究¹¹⁾がある。指し手のレーティングや killer heuristics など様々な工夫が行われているが、佐藤らによれば実力は初段程度に達することはできていない。評価関数との関連では、静止探索の利用がゲーム木探索において行われている。これは、互角の局面において駒損する手を選ばないために用いられた手法で、実際に有効であることが実験から示されている。

3. 提案手法

モンテカルロ木探索の問題点として、優勢な局面では、どの手を選んででも勝ちになりやすく、良くない指し手を選択する傾向があることと、囲碁におけるシチョウ、将棋における詰みや駒の取り合いなど探索が必要となるような局面に弱いことが挙げられる。特に後者は、将棋において駒の取り合いが多く発生する上に勝敗に直結しやすいため、大きな問題点となる。

前者の問題点を解決するため、評価値を利用する時にルート局面の評価値との差として利用することを提案する。例えば、探索末端において評価値の正負を勝敗としていた手法は、探索末端の評価値とルート局面の評価値の大小を勝敗とすることになる。これにより、ルート局面において優勢であってもより良くなる手を選ぶようになることが期待される。

次に、後者の問題点の解決のために、評価値を得るさいに静止探索を用いることを提案する。静止探索とは、駒の取り合いなどを探索することで探索末端における評価値を安定させるものであり、将棋においては一般に静止探索が利用されている。

従来手法と先行研究の手法、提案手法の違いは、プレイアウトにおける勝敗の決め方であり、次のようになっている。

- 従来手法 (モンテカルロ木探索): プレイアウトの勝敗
- 先行研究 1 (EF-Leaf): プレイアウトが一定の深さに達した時の評価値の正負
- 先行研究 2 (EF-Cutoff): プレイアウトにおいて、評価値が閾値を超えた (下回った) なら勝 (負)
- 提案手法 (QS-Leaf-Diff): プレイアウトが一定の深さに達した時の静止探索による評価値とルート局面の評価値との大小
- 提案手法 (QS-Cutoff-Diff): プレイアウトにおいて、静止探索による評価値が、ルート局面の評価値を十分に超えた (下回った) なら勝 (負)

なお、各手法のパラメータとして、EF-Leaf と QS-Leaf-Diff ではプレイアウトを打ち切る深さ、EF-Cutoff と QS-Cutoff-Diff では評価値の閾値がある。これらのパラメータは問題集の実験結果から決定した。

4. 実装

本研究におけるゲーム木探索とプレイアウトそれぞれの実装について以下で説明する。各パラメータについては、対戦や問題集などの予備実験から決定した。

4.1 ゲーム木探索

手の選択、展開には Upper Confidence bound for Tree を基本とし、Progressive Bias¹⁾, First Play Urgency (FPU) などの手法を利用した。最終的には、ノードを展開する閾値を 1, Progressive Bias の係数を 20, FPU の値を 0.9 とした。

また、ゲーム木探索において詰みを発見できるように、1 手詰判定関数を利用している他、Monte-Carlo Tree Search Solver⁹⁾ を実装した。ゲーム木探索において、詰みによる勝敗を得た場合に、特別な勝ち負けの値を使い、その値については min-max 的に利用することで詰みを発見する手法である。

4.2 プレイアウト

勝敗の判定については、1 手詰判定関数を利用した。プレイアウト中の指し手の選択は以下のように 2 つの手法を利用した。評価関数を使うモンテカルロ木探索では、レーティング上位 5 手からランダムに選択し、評価関数を使わないモンテカルロ木探索では以下の順に選択した。

- 取り返しの手があれば、その中からランダムに選ぶ
- 駒取りの手があれば、その中からランダムに選ぶ
- 残りの手からランダムに選ぶ

表 1 問題集の解答結果

プログラム	正答数	パラメータ
モンテカルロ木探索	136	
EF-Leaf	186	深さ: 6
EF-Cutoff	223	閾値: 歩 2 枚
QS-Leaf-Diff	256	深さ: 2
QS-Cutoff-Diff	274	閾値: 歩 2 枚

表 2 自己対戦結果

		勝	分	敗
EF-Leaf	モンテカルロ木探索	26	5	17
EF-Cutoff	モンテカルロ木探索	23	1	24
EF-Cutoff	EF-Leaf	17	1	30
QS-Leaf-Diff	EF-Leaf	43	0	5
QS-Cutoff-Diff	EF-Cutoff	41	0	7
QS-Cutoff-Diff	QS-Leaf-Diff	23	1	24

5. 実 験

提案手法の有効性を示すために、問題集の正答数の比較、自己対戦、レーティングの計測を行なった。プログラムの実装には、GPS 将棋^{*1}(ver.2411 Open Shogi Library^{*2} ver.4208) を利用し、実験環境として CPU Core2Quad Q9650 3.00GHz, メモリ 8GB のマシンを利用した。なお、並列化は行っていない。

5.1 問題集

プログラムの性能を評価するため、問題集「ラクラク次の一手」¹³⁾¹⁴⁾ (全 432 問) を 1 問 10 秒で解かせ、正答数を比較した。なお、複数のパラメータで実験を行い、正答数が最も多かったものを掲載している。実験結果を表 1 にまとめた。

実験結果から、評価関数の利用により性能の改善が確認でき、提案手法によりさらに性能が改善していることがわかる。提案手法の組み合わせである QS-Cutoff-Diff がもっとも有効であると考えられる。

5.2 自己対戦

続いて、1 手 10 秒の対戦で 48 局の自己対戦を行った結果を表 2 にまとめる。パラメータについては問題集の実験で決めたパラメータを用いた。定跡や戦型の偏りを避けるため、プロの棋譜から序盤の局面を選び対局を開始させた。

先行研究の手法は、評価関数を利用しない手法に対して有意に勝ち越すことはできなかったが、提案手法は先行研究の手法に有意に勝ち越しており、提案手法の有効性を示すことができた。一方、問題集の結果と異なる結果として、評価関数を利用しないモンテカルロ木探索が評価関数を利用するプログラム (EF-Leaf, EF-Cutoff) と同等の強さとなっていることと EF-Leaf が EF-Cutoff に比べて成績が良いことが挙げられる。

5.3 レーティング

提案手法を採用したモンテカルロ木探索将棋の性能を測るため、コンピュータ将棋連続対局場所である

表 3 floodgate におけるレーティング

プログラム	レーティング
gps_l	2464
gps_normal	2150
gps_500	1540
提案手法	1821

表 4 各手法による正答数の増減

	EF	+Diff	+QS	+QS-Diff
Leaf	186	+52	+43	+70
Cutoff	223	-5	+17	+52

floodgate^{*3} で対戦を行わせ、レーティングを計測した。問題集の正答数が最も多かった QS-Cutoff-Diff を採用し、mcts_qcdem という名前で参加した。2010/9/12 時点でのレーティングは 1821 (2 週間レーティング) であった。他の GPS 将棋のレーティングを表 3 にまとめた。なお、gps_l は通常の GPS 将棋で、gps_normal は手調整の評価関数を利用した古いバージョンで (レーティングは固定)、gps_500 は非常に浅い探索のみの GPS 将棋となっている。

提案手法を用いたモンテカルロ木探索はアルファベータ木探索を上回ることができていない。

5.4 個々の手法の効果

本研究では、ルート局面の評価値の差を見る手法 (Diff) と評価値を求める際に静止探索を利用する手法 (QS) を提案し、これまでの実験から、両者の利用による性能改善を確認してきた。ここでは、それぞれの貢献を見るために、個別に適用した場合の性能評価を行った。

それぞれを実装し、5.1 節と同じ条件で問題集を解かせた結果を表 4 にまとめた。表中の +/- のついた数字は EF-Leaf, EF-Cutoff との正答数の差を表している。探索末端において評価値を利用する手法 (Leaf) では、それぞれの工夫により大きな性能向上が得られている。一方、評価値が閾値を超えるかで判定する手法 (Cutoff) では、単独では Diff は効果が得られていない。

*1 <http://gps.tanaka.ecc.u-tokyo.ac.jp/gpsshogi/>

*2 <http://gps.tanaka.ecc.u-tokyo.ac.jp/gpsshogi/pukiwiki.php?OpenShogiLib>

*3 <http://wdoor.c.u-tokyo.ac.jp/shogi/floodgate.html>

6. おわりに

評価関数を利用したモンテカルロ木探索を将棋へと適用した。優勢な局面での緩みや駒取りのような探索が必要となる局面といった従来のモンテカルロ木探索における問題点と性能の改善のために、評価値を利用する時にはルート局面の評価値との差を利用し、評価値を得る際に静止探索を利用するという手法を提案した。実験結果から提案手法による性能改善が確認でき、有効性を示すことができた。

今後は、並列化や先行研究で用いられる手法の適用など、さらなる改良を行っていく。特に、佐藤らの研究¹¹⁾で用いられているゲーム木探索における静止探索の利用について、提案手法との併用による効果の確認を行いたい。また、静止探索の利用によって性能改善が得られたが、詰みは勝敗に直結しており、詰将棋探索を利用することで性能改善が期待できる。

参考文献

- 1) Guillaume M. J-B. Chaslot, Mark H.M. Winands, H.Jaap VanDen Herik, Jos W. H.M. Uiterwijk, and Bruno Bouzy. Progressive strategies for monte-carlo tree search. *New Mathematics and Natural Computation (NMNC)*, Vol.4, No.03, pp. 343–357, 2008.
- 2) Rémi Coulom. Computing elo ratings of move patterns in the game of go. In *Computer Games Workshop*, Amsterdam / The Netherlands, 2007.
- 3) Hilmar Finnsson and Yngvi Björnsson. Simulation-based approach to general game playing. In *Proceedings of the Twenty-Third National Conference on Artificial Intelligence (AAAI-2008)*, pp. 259–264, 2008.
- 4) Sylvain Gelly and David Silver. Combining online and offline knowledge in uct. In *ICML '07: Proceedings of the 24th international conference on Machine learning*, pp. 273–280, New York, NY, USA, 2007. ACM.
- 5) Julien Kloetzer, Hiroyuki Iida, and Bruno Bouzy. The monte-carlo approach in amazons. In *Computer Games Workshop*, 2007.
- 6) Tomáš Kozelek. Methods of mcts and the game arimaa. Master's thesis, Charles University in Prague, 2009.
- 7) RichardJ. Lorentz. Amazons discover monte-carlo. In *CG '08: Proceedings of the 6th international conference on Computers and Games*, pp. 13–24, Berlin, Heidelberg, 2008. Springer-Verlag.
- 8) MarkH. Winands and Yngvi Björnsson. Evaluation function based monte-carlo loa. In *Advances in Computer Games*, Vol. 6048 of *Lecture Notes in Computer Science*, pp. 33–44. Springer, 2009.
- 9) Mark H. Winands, Yngvi Björnsson, and Jahn-Takeshi Saito. Monte-carlo tree search solver. In *CG '08: Proceedings of the 6th international conference on Computers and Games*, pp. 25–36, Berlin, Heidelberg, 2008. Springer-Verlag.
- 10) 美添一樹. モンテカルロ木探索 – コンピュータ囲碁に革命を起こした新手法. *情報処理*, Vol.49, No.6, pp. 686–693, 2008.
- 11) 佐藤佳州, 高橋大介. モンテカルロ木探索によるコンピュータ将棋. 第 13 回ゲームプログラミングワークショップ, pp. 1–8, 2008.
- 12) 金子知適, 山口和紀. 将棋の棋譜を利用した, 大規模な評価関数の調整. 第 13 回ゲームプログラミングワークショップ, pp. 152–159, November 2008.
- 13) 日本将棋連盟書籍 (編). ラクラク次の一手 基本手筋集. 日本将棋連盟, 2003.
- 14) 日本将棋連盟書籍 (編). ラクラク次の一手 2 基本手筋集. 日本将棋連盟, 2003.
- 15) 保木邦仁. 局面評価の学習を旨とした探索結果の最適制御. 第 11 回ゲームプログラミングワークショップ, pp. 78–83, 2006.