C-02

花札の戦略決定 -UCT 探索の性能検証-

北海道科学大学 ○髙岡 勇樹 川上 敬 大江 亮介 三田村 保 木下 正博

要旨
近年，ゲーム木探索に UCT(UCB Applied to Tree)探索を用いることが主流になっている．この UCT 探索を用いることでゲーム木を効率的に探索することが可能である．本報告では，UCT 探索を花札の意思決定アルゴリズムに導入し，人との対戦実験を行って UCT 探索の性能を検証する．

1. はじめに

不完全情報ゲームにおいては，盤面に隠されている情報が存在し，全ての情報が公開されている完全情報ゲームで有効とされている従来の手法を戦略決定に使用するのは難しい場合がある．この困難を発生させる理由としては，隠されている情報を推定しなければならない点や，とりうる全ての状態を探索対象とする点，状態の好ましさを表現する評価関数を作成しにくい点などが挙げられる．

このような問題を解決するための方法として，近年コンピュータ囲碁やコンピュータ将棋などの分野で注目されている UCT(UCB applied to Tree)[1]を利用することが考えられる．UCT とは，モンテカルロ探索法を木探索に応用した手法で，シミュレーションの結果を勘案して好ましいと思われる手を重点的に探索する方法である．この UCT を不完全情報ゲームに適用し，UCT が不完全情報ゲームに有効であるとした例[2]も報告されているが，他のゲームへの適用例はあまり見られない．

そこで本研究では，不完全情報ゲームの一種である花札の戦略決定の為の手法として，UCT を利用した最善手決定アルゴリズムを提案する．この手法を用いることによって広大な探索空間を効率よく探索することができ，かつ評価関数を必要としないため，あらゆる状況に対応ができると考えられる．

ここでの実験としては，UCT 探索で手を決定するコンピュータプレイヤを作成し，人間が相手となって花札の対局を行う．UCT プレイヤとはゲームの各局面で切る札の選択決定に UCT を適用し，どの札を切れば高い報酬が得られるかを探索しながらゲームを進行させるプレイヤである．両者の対局ごとの得点収支を観測し，平均獲得報酬を分析することを実験とする．

2. UCT 探索

ここでは，本研究で採用した UCT(UCB applied to Tree)および UCT の基になったモンテカルロ法について説明する．

2.1 モンテカルロ法

モンテカルロ法は一般的に多数の乱数を用いてシミュレーションや数値計算を行うことによって確率的に結果を求める手法であるが，これをゲームに適用する場合，乱数を用いてゲームを進行させ，その結果を蓄積することで最適手を求める手法が多くとられる．具体的には各種ゲームで規定のルールに従って終局までランダムにプレイし(以下プレイアウトと呼ぶ)，ゲームのスコアを計算し，それを局面の評価とする．この手法は局面の評価に評価関数を必要としないので，評価関数の設計が困難であるとされていたコンピュータ囲碁やコンピュータ将棋で注目されてきた．しかし欠点として，

・　乱数を用いるために不確実性が高く結局は相手のミスを期待することになる
・　プレイアウト回数を大きくしても一定回数以上では実力が変わらない

といった点がある[3]

このモンテカルロ法を木探索に組み込んだのがモンテカルロ木探索である．モンテカルロ木探索の特徴は，図 1 のように有望な手に多くのプレイアウト回数を割り当てること，また図 2 のようにゲーム木が成長することである．ここでは黒のプレイヤに対して木探索を行っている．

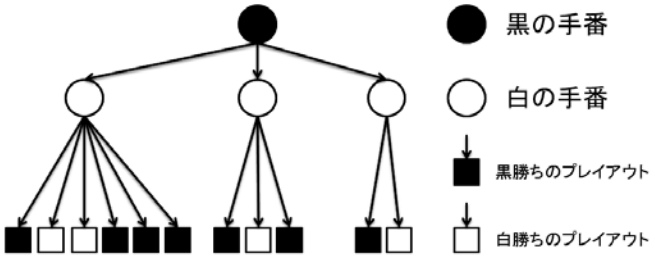

Fig. 1 Monte Calro Tree Search

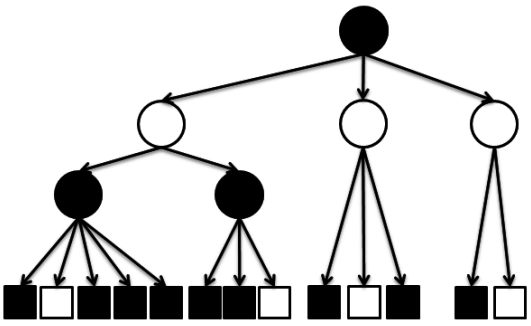

Fig. 2 Monte Calro Tree Search

このモンテカルロ木探索により，囲碁プログラムの CrazyStone は同じく囲碁プログラムの GNU Go に対して勝率 6 割台にまでなった[3].

2.2 UCT

UCT とは，多腕バンディット問題(Multi-Armed Bandit)を解決するために提案された UCB1(Upper Confidence bounds)をモンテカルロ木探索に応用したものである．多腕バンディット問題とは，腕が複数存在するスロットマシンのイメージであり，各腕がそれぞれの確率でコインが払い出され，どの腕にコインを投入するべきか考えるものである．これを言い換えると，たくさんあるスロットマシンのうち，どこにどれだけのコインを投入すれば儲けが出るか，となる．

この問題を解くために提案されたのが UCB1 で，2002 年

に Auer らによって発表された[4]．この戦略はスロットマシン毎に UCB という値を計算し，最も値が高いマシンにコインを投入していくもので，計算量を抑えながら高い払い出しを受けられる戦略である．UCB1 では次式により各スロットマシンの評価値を算出する．

$$UCB(i) = \bar{X}_i + c\sqrt{\frac{2\ln n}{n_i}} \qquad (1)$$

式(1)中の$\bar{X}_i$は番目のマシンのその局面での報酬期待値，n_iはi番目のマシンのコイン投入数，nは全てのマシンのコイン投入数の合計である．cは定数だが，通常は$c = 1$が用いられる．

この式により，期待値が高いマシンほど選択されやすくなり，期待値は低いが選択回数が少ないマシンは第 2 項によって考慮することができる．この手法を木探索に適用したのが UCT である．

UCT ではルートノードより UCB 値が最も高い子ノードを順にたどり，末端のノードに到着したらプレイアウトを行う．その結果によってたどったノードの UCB 値を更新し，再び探索を行う．これを繰り返し，末端の子ノードの探索回数が閾値を超えた場合はノードを展開する(図 2)．以上を規定回数行い，UCB の値が最も高い手を次の手とする．

UCT はシミュレーションを繰り返して最適解を得るアルゴリズムであるので，質の良い評価関数の設計が不要であるのも特徴である．

3. UCT を用いた花札の戦略決定

ここでは，本研究で行ったシミュレーション実験に関して述べる．

3.1 花札について

本研究では花札の競技方法のうち，代表的な二人打ちルールである「こいこい」を対象問題とした．以下，花札の概要とこいこいのルールを説明する．

花札とは，札の枚数が 48 枚で，1 月から 12 月の 12 の月に別れている．各月に 4 枚の札があり，同じ月の札と合わせると札を取ることができる．

こいこいとは，親と子に別れたプレイヤが札を取り合い，規定の役を完成させることを目指すゲームである．役は十数種類あり，それぞれに役の価値(役代と呼ぶ)が定められており，単位は文である．自分の手番では，手札より 1 枚札を切り，取れる札があるならば自分の持ち札とする．続いて山札より 1 枚引き，同じようにする．これを繰り返し役が完成した場合プレイヤは，そこで勝負を続行しさらなる役を作りに行くか，勝負を終えて役代を相手より得るかを選択することができる．この一連の流れを 12 回繰り返し，文数の多いプレイヤが勝利する．

図 3 は役の一例である．例として赤短の役を挙げた．赤短は，松・梅・桜の短冊札から成り立つ．このように，こいこいの役は特定の札を集めることによって成立する．

Fig. 3 One of the winning hands of Koi-Koi(Aka-Tan)

3.2 実験内容

UCT に関する実験として，UCT 探索を行うプレイヤを作成し，人間との対局を行った．実験は初心者から花札が得意な人間まで，幅広いレベルのプレイヤで対戦をした．また通常行われる花札の対局と同様，12 回のゲームで 1 対局とする．

実験環境は Core i5 2.8GHz，メモリ 4GB で行った．式(1)中のパラメータは$c = 1$，報酬は獲得文数とし，ループ回数は 2,000 回とした．またゲーム木探索中の 1 エピソードはどちらかが役を完成させるか手札を出し終えるまでを行い，1 エピソードの終了を 1 回の UCT として，規定のループ回数 UCT を行った後に最も有望な札を切るものと設定した．なお，現在は木を成長させずに，1 段読みのみでシミュレーションを行っている．すなわち，図 1 の木を図 2 のように成長させる手順は行わないものとする．

3.3 実験結果

20 対局の結果，人間対 UCT プレイヤの対局は「UCT プレイヤ側の 7.00 文勝ち」という結果になった．花札はゼロサムゲームであるので，人間側は同じ文数だけ負けているということになる．これにより，UCT プレイヤの実力が高いことを示している．得られた結果から，UCT を用いることで既存の知識が無くても，各局面で有効な手を求められていることがわかる．

4. 考察と今後の課題

本研究では不完全情報ゲームの 1 つである花札の手を探索する手法として，UCT を利用した探索手法を提案した．行った実験では，UCT プレイヤが人間に平均で勝ち越しており，UCT が花札に有効である可能性があることがわかった．

今後の課題としては，実験回数の増加，アルゴリズムの改善などが挙げられる．行った実験回数が 20 対局と少ないので，より多くの実験を行う必要がある．また，実験を行う人間の選定も必要になってくる．今回は初心者も対象としているため，初心者が負けた分の文数が UCT 側の勝ちに繋がっている場合もある．よって，対局を行う人間の検討もしていきたい．

アルゴリズムの改善については，木の成長をさせることが必要である．現状では木探索を行うときに木を成長させていないが，本来は木が成長していくものである．そのため，現段階では UCT の初歩にあり，木を成長させて実験を行いたいと考える．また，UCT のループ回数が 2,000 回と少ないので，ループ回数も増やす必要がある．

参考文献

[1] L. Kocsis and C. Szepesvari : Bandit based monte-carlo planning , European Conference on Machine Learning , 282/293(2006)

[2] J. Schäfer, M. Buro, and K. Hartmann : The UCT Algorithm Applied to Games with Imperfect Information , Diploma thesis , Otto-von-Guericke-Universität Magdeburg(2008)

[3] 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, 686/693 (2008)

[4] Auer, P., Cesa-Bianchi, N. and Fischer, P. : Finite-time Analysis of the Multiarmed Bandit Problem , Machine Learning, Vol.47, 235/256 (2002)