「芝浦将棋 Softmax」のチーム紹介

2017 年 3 月 14 日
芝浦工業大学情報工学科
五十嵐治一，原 悠一

1. はじめに

本稿は，第 27 回世界コンピュータ将棋選手権（2017 年 5 月 3 日～5 日開催）に出場予定の「芝浦将棋 Softmax」（シバウラショウギ　ソフトマックス）のアピール文書です．本チームは「芝浦将棋 Jr.」から分離した初参加のチームです．探索手法が従来の Min-max 探索（α β探索）とは異なる Softmax 探索である点が大きく異なります．ただし，合法手生成までは芝浦将棋 Jr.と共通で，選手権公式ライブラリとして登録されている「芝浦将棋 Jr.合法手生成プログラム」[1]を使用しています．棋力的には最新版の芝浦将棋 Jr.には及びませんが，アルゴリズムが単純でコーディングの容易さや，並列性に優れています．以下，簡単に本チームの特徴を紹介していきます.

２．開発メンバー

五十嵐は芝浦工業大学工学部情報工学科に勤務する教員です．原は五十嵐研究室に所属し，修士課程において Softmax 探索を用いたコンピュータ将棋プログラムの開発と研究を行いました（2017 年 3 月修了）．

3.「芝浦将棋 Softmax」の特徴

本チームの探索方式の考案には，文献[2]の研究が基になっています．本チームの特徴を，以下の１）～５）のようにまとめました.

１）芝浦将棋 Jr. の合法手生成ルーチンを使用

芝浦将棋 Jr. では盤面表現のデータ構造を独自の Magic bitboard を用いて駒の利き場所での駒の配置状況などを計算しています[3]．この計算を含む合法手生成のプログラムは「芝浦将棋 Jr.合法手生成プログラム」の名称で選手権公式ライブラリとして登録されています．芝浦将棋 Softmax はこの合法手生成プログラムをそのまま使用しています.

２）Softmax 探索を使用

現在のチェスや将棋のプログラムは「Min-max 探索」という探索方式をほぼ 100%採用

しています．これには探索木のすべてのノードを探索する必要がありますが，α β カットなどの枝刈りの処理により探索にかかる計算時間を短縮しています．これに対して，探索するノードを限定したり，優先順位をつけて選択的に探索する「選択探索」という探索方式があります．本チームはノードの選択方式としてノード評価値の min-max 演算ではなく，確率分布に基づく選択（Softmax 探索）を使用しています．

３）ノードの評価関数を用いたボルツマン分布による確率的なノード選択

Softmax 探索には指し手の良さを用いたボルツマン分布を利用します．すなわち，各ノードでの指し手の選択確率を次の式で計算し，その確率に従ってノードを選択していきます．

$$\pi(a|s) = exp(E_a(a;s)/T)/\sum_{x\in A(s)} exp(E_a(x;s)/T) \qquad (1)$$

ただし，s は局面（ノード），a は指し手，$E_a(a;s)$は局面 s における指し手 a の良さですが，指した後の局面ノードの評価関数 $E_s(s)$で置き換えることにします．$A(s)$は s における合法手の集合，T は温度と呼ばれているパラメータです．温度が低ければ最良優先探索に，温度が高ければランダム探索に近づきます．ノードの評価関数は，探索木の末端ノード（leaf）であればそのノードの局面評価関数で定義します．一方，内部ノードであれば子ノード $v(x;s)$ の評価値 $E_s(v)$をその子ノードの選択確率 $\pi(x|s)$で重みづけた期待値

$$E_s(s) = \sum_{x\in A(s)} \pi(x|s)E_s\big(v(x;s)\big) \qquad (2)$$

で定義します．

４）深さ制御とバックアップ操作

探索の全体の流れを図１に示します．ルートノードから，３）の選択法に従ってノードを選択し，末端ノードまで到達すると，一段階だけ探索木を展開します．展開後は新たな末端ノードの評価値を局面評価関数で計算し，その値をルートノードへ向けて(2)の計算を繰り返し，ルートノードまでの経路上のノード評価値を更新していきます．我々はこの更新操作を「バックアップ操作」と呼び，本探索方式を“Randomized Softmax Search”または“Monte Carlo Softmax Serach”と称しています．名前の由来は，ルートノードから末端ノードへ到達するまで，(1)の選択確率に従って確率的にノード選択を行って経路が生成される過程は，指し手の良さ（＝その手の子ノードの評価値の期待値）を求めるためのモンテカルロ・サンプリングに相当するからです．

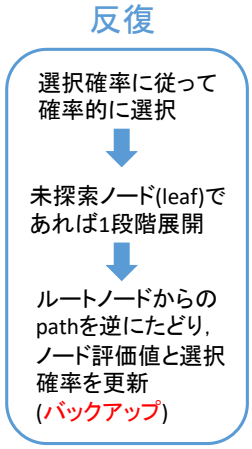

図１

上記のモンテカルロ・サンプリングを一定回数あるいは一定時間行った後，確率値の最も高い子ノードを選択し（greedy 選択：softmax 選択で T→0 とした場合に相当），得られた手順を最善応手手順と決定します．

今回のバージョンでは深さ制御のために特別な処理を何もやっておりません．しかし，(1)の選択確率の値を用いて，決定論的な最良優先探索を行う探索法も考えられます．これは選択確率の積を実現確率と定義し，実現確率の閾値を深さとする反復深化法と結びつけることができます[4]．この方式も実装して対局実験を少しだけ行ってみたのですが，上で述べた選択確率に基づく確率的なサンプリング方式の方が棋力が上で，並列処理の効果も高かったので今回採用しました．

５）評価関数について

現在のところ，評価関数の特徴量は，選手権公式ライブラリである Bonanza (Ver. 6.0.0) [5] のものをそのまま使用しています．しかし，３）で述べたモンテカルロ・サンプリングによる探索方式の下で生成された探索木の全 leaf に出現する特徴量の重みを同様なモンテカルロ・サンプリングとバックアップ操作だけで学習することが可能です（未発表）．将来的にはこの学習法も実装していく予定です．

なお，末端ノードでの局面評価には静止探索（駒の取り合いだけを考慮する探索）を行って，その結果を局面評価として返す処理を行っています．現バージョンのプログラムでは，この静止探索においては高速化のために従来の α β 探索を使用しています．

4. 棋力について

2015 年 11 月に開催された「第 3 回将棋電王トーナメント」へ出場したバージョンの芝浦将棋 Jr. と対局実験を行ってみました．このときの芝浦将棋 Jr. は探索の並列化に対応しておらず，参加 28 チーム中 16 位の成績でした．1 手 10 秒，300 局，先後入れ替え，Ponder なしの条件で実験しました．結果を表 1 に記します．

表 1　芝浦将棋 Jr. (2015 年 11 月時点のバージョン) との対局実験結果

スレッド数	勝　負　分	勝率
1	96　196　38	32.8%
6	188　109　3	63.3%

なお，上記の 6 スレッドを使用した対局実験では，平均最大探索深さは 15.5 手，平均最善応手手順の深さは 9.5 手でした．

5. 今後の課題

実は 12 スレッドでも対局実験を少しやりましたが，棋力や読みの深さが大きく変化する

ことはありませんでした．また，簡単な詰みを逃したり，駒のやり取りを伴わない場合に相手の罠に簡単にはまったりするなどの欠点がときどき見受けられました．どのような形の探索木が出来上がっているのかを解析して，最適な温度パラメータを見つけることや，探索中に適応的に温度パラメータを調整するなどの何らかの深さ制御が必要と考えています．また，静止探索には従来のαβカットを使用していますが，静止探索も softmax 探索だけで計算することを検討中です．

さらに，探索と同様に局面評価関数の学習をモンテカルロ・サンプリングによって計算することが可能です．すでに softmax 探索における方策勾配の計算式を文献[2]において提案していましたが，最近，生成後の探索木を用いてモンテカルロ・サンプリングを行うことにより数値的に厳密に計算できることが分かりました．これにより，探索木のすべての末端ノードに出現した局面に含まれる特徴量の重み値の更新が可能になります．局面評価関数の学習法としては Bonanza メソッド[5]が有名ですが，この学習法ではルートノードにおける各合法手から伸びる最善応手手順で得られた末端ノード局面に含まれる特徴量重みだけしか更新されません．したがって，本学習法であれば従来よりは少ない対局数で学習が可能になることが期待されます．

6. おわりに

現在のコンピュータ将棋プログラムは Stockfish[6]などのチェスプログラムの探索方式をソースコードのレベルで大きな影響を受けています．それに対して，本チームが考案，開発しているような Softmax 探索をベースにした方式は今の将棋プログラムではほとんど試みられておりません．本探索方式はモンテカルロ木探索の一種と思われますが，プレイアウトを行わない点が囲碁プログラムとも異なります．本探索方式は，プログラム作成が容易で，並列化の効果も高い上に，他のゲームプログラムへの適用も容易であるという点で汎用性にも優れていると考えています．

まだまだ問題点も多いのですが，新しい探索方式と学習方式を研究する上では面白さが多く，開発者自身，今後の展開を楽しみにしております．最終的には，プロ棋士の棋譜を用いることなく，コンピュータ自身が自己対局を（あるいは他者との他流試合も）通して，探索法や局面評価関数を学習し，人類の棋力を超えて，新しい定跡や戦法を創出し，棋士や将棋ファンを大いに楽しませてくれることを目標としております．

参考文献

[1] 「芝浦将棋 Jr.合法手生成プログラム」の機能説明書とプログラムは次のページからダウンロードできます：http://www2.computer-shogi.org/library/

[2] 五十嵐治一，森岡祐一，山本一将，“方策勾配法による静的局面評価関数の強化学習についての一考察”，第 17 回ゲームプログラミングワークショップ(GPW2012)予稿集，pp.118-121 (2012).

[3] 例えば，http://www2.computer-shogi.org/wcsc26/appeal/Shibaura_Shougi_Jr./appeal.pdf に記載されています.
[4] 原悠一，五十嵐治一，森岡祐一，山本一将，“ソフトマックス戦略と実現確率による深さ制御を用いたシンプルなゲーム木探索方式”，第 21 回ゲーム・プログラミング・ワークショップ(GPW2016)2016 予稿集，pp.108-111(2016).
[5] Bonanza のホームページ，http://www.geocities.jp/bonanza_shogi/
[6] Stockfish のホームページ，https://stockfishchess.org/